**PCT** WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

WO 88/02024

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 4 : C12N 15/00, 9/88

A1

(11) International Publication Number: WO 88/ 0202

(43) International Publication Date: 24 March 1988 (24.03.8

(21) International Application Number: PCT/GB87/00628

(22) International Filing Date: 8 September 1987 (08.09.87)

(31) Priority Application Number: 8621626

(32) Priority Date: 8 September 1986 (08.09.86)

(33) Priority Country: GB

(71) Applicant *(for all designated States except US)*: THE PUBLIC HEALTH LABORATORY SERVICE BOARD [GB/GB]; 61 Colindale Avenue, London NW9 5EQ (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)* : ANSON, John [GB/GB]; 3 Avon Terrace, Salisbury, Wiltshire (GB). GILBERT, Harold [GB/GB]; 16 Kells Gardens, Low Fell, Gateshead, Tyne and Wear (GB). ORAM, Jon [GB/GB]; Candle Cottage, Wick Lane, Manningford, Bohune, Pewsey, Wiltshire (GB). MINTON, Nigel, Peter [GB/GB]; Thatch End, 16 Newton Tohey, Salisbury, Wiltshire (GB).

(74) Agent: BECKHAM, R., W.; Ministry of Defenc Procument Executive, Patents 1A(4), Room 201 Empress State Building, Lillie Road, London SW 1TR (GB).

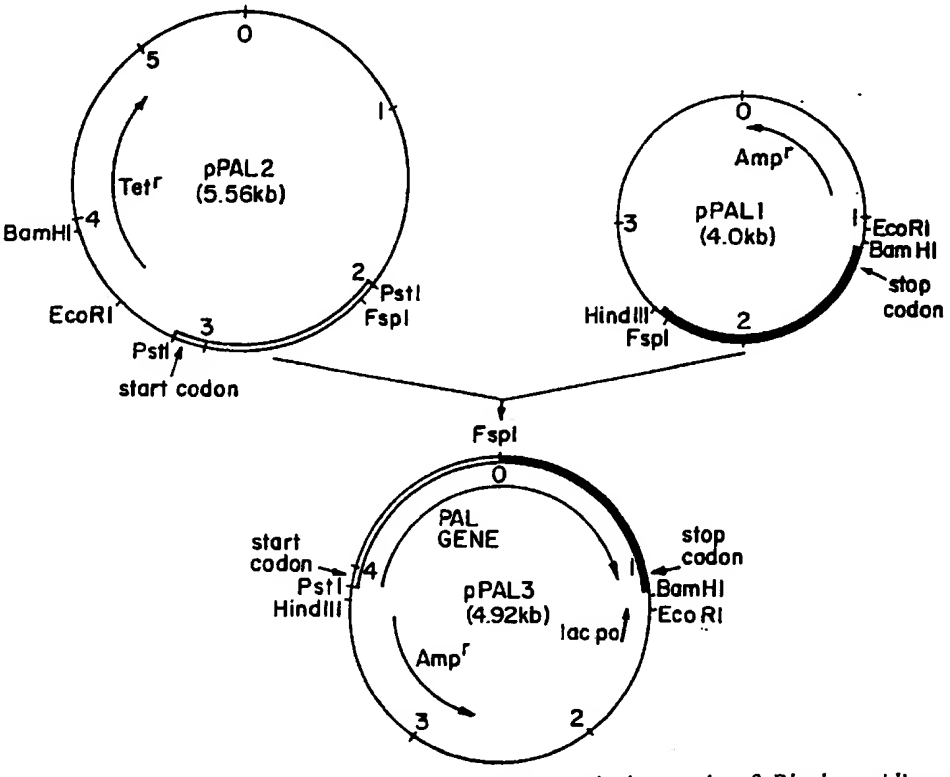
(81) Designated States: AT (European patent), BE (European patent), CH (European patent), DE (European patent), DK, FR (European patent), GB, GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), NO, SE (European patent), US.

**Published**

*With international search report.*

*Before the expiration of the time limit for amending t claims and to be republished in the event of the recei of amendments.*

(54) Title: PRODUCTION OF PHENYLALANINE AMMONIA LYASE




(57) Abstract

For use in genetic engineering a gene is provided, derived from a PAL-producing strain of *Rhodosporidium toroloides*, from which non coding introns have been excised. The gene may be inserted into plasmid vectors which may be introduced into heterologous organisms so that PAL is expressed. A method of preparing the gene is provided, and its polynucleotide sequence is listed.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | |
|---|---|---|---|---|---|
| **AT** | Austria | **FR** | France | **ML** | Mali |
| **AU** | Australia | **GA** | Gabon | **MR** | Mauritania |
| **BB** | Barbados | **GB** | United Kingdom | **MW** | Malawi |
| **BE** | Belgium | **HU** | Hungary | **NL** | Netherlands |
| **BG** | Bulgaria | **IT** | Italy | **NO** | Norway |
| **BJ** | Benin | **JP** | Japan | **RO** | Romania |
| **BR** | Brazil | **KP** | Democratic People's Republic of Korea | **SD** | Sudan |
| **CF** | Central African Republic | | | **SE** | Sweden |
| **CG** | Congo | **KR** | Republic of Korea | **SN** | Senegal |
| **CH** | Switzerland | **LI** | Liechtenstein | **SU** | Soviet Union |
| **CM** | Cameroon | **LK** | Sri Lanka | **TD** | Chad |
| **DE** | Germany, Federal Republic of | **LU** | Luxembourg | **TG** | Togo |
| **DK** | Denmark | **MC** | Monaco | **US** | United States of America |
| **FI** | Finland | **MG** | Madagascar | | |

# PRODUCTION OF PHENYLALANINE AMMONIA LYASE

This invention relates to genetic material which encodes the protein phenylalanine ammonia lyase (herein abbreviated to 'PAL') and in particular to such genetic material which lacks the intervening non-coding DNA (introns) normally found in the PAL - encoding gene in its natural state.

Phenylalanine ammonialyase (PAL; EC 4.3.1.5) which occurs in plants, yeasts, fungi, and streptomycetes catalyzes the nonoxidative deamination of L-phenylalanine to trans-cinnamic acid (see Gilbert et al., 1985). The enzyme has a potential role in the treatment and diagnosis of phenylketonuria (Ambrus et al., 1978) and has industrial applications in the synthesis of L-phenylalanine from trans-cinnamic acid (Yamada et al., 1981). In plants the enzyme, involved in flavanoid biosynthesis, is induced by illumination while in gherkin and mustard seedlings induction is the result of activation of a constitutive pool of inactive enzyme (Attridge et al., 1974). Illumination elicits de novo synthesis of the enzyme in other botanical species (Schroder et al., 1979). Gherkin, apple, sweet potatoe, and sunflower PAL is also regulated by a specific inactivating system (Tan, 1980).

In some basidiomycete yeast phenylalanine can act as sole source of carbon, nitrogen, and energy. As PAL catalyzes the initial reaction in the catabolism of the amino acid, the enzyme plays a key role in regulating phenylalanine metabolism. In Rhodosporidium toruloides PAL is induced by the presence of L-phenylalanine or L-tyrosine (Marusich et al., 1981). Glucose, and ammonia in the presence of glucose, repress PAL synthesis (Marusich et al., 1981), while induction of PAL activity is the result of de novo synthesis of the enzyme rather than activation of an inactive precursor or a decrease in the rate of PAL degradation (Gilbert and Tully, 1982). Glucose represses PAL synthesis but has no effect upon stability of the enzyme, whereas ammonia prevents uptake of phenylalanine and so may repress enzyme synthesis through inducer exclusion (Gilbert and Tully, 1982). In vitro translation data of mRNA, isolated from R. toruloides grown under different physiological conditions, showed that

phenylalanine, ammonia and glucose regulate PAL synthesis by adjusting the level of functional PAL mRNA (Gilbert et al., 1983).

In recent years genetic engineering methods have been developed whereby microorganisms which are common or which can easily be grown on an industrial scale, in particular certain bacteria or yeasts, have their genetic material (DNA sequences) modified so that they produce a desired compound eg a protein. Broadly this is achieved by inserting into the host microorganism a plasmid consisting of a gene which is a polynucleotide sequence which encodesthe compound, together with other genetic material which instructs the host's genetic apparatus to synthesise the compound.

The gene encoding PAL has recently been cloned as a 8.5 kb genomic PstI fragment (Gilbert et al., 1985). These studies indicated that PAL is synthesised from a monocistronic mRNA of 2.5 kb, and that the gene is present as a single copy in the the R. toruloides genome. The introduction of the cloned PAL gene into both E. coli (Gilbert et al., 1985) and Saccharomyces cerevisae (Tully and Gilbert, 1985) did not result in the production of PAL protein.

Although attempts have been made along these lines to introduce the cloned PAL - encoding gene from R-toruloides into the microorganism E-coli (Gilbert et al; 1985) and into the yeast Saccharomyces Cerevisae (Tully and Gilbert, 1985), these heterologous hosts did not then produce PAL protein.

It is an object of the invention to provide genetic material which may be introduced into host organisms other than R-toruloides, which hosts will then produce PAL protein. Other objects and advantages of the invention will be apparent from the following description.

According to a first aspect of the invention there is provided an intron-free structural gene, derived from a corresponding intron-containing structural gene from a eukaryotic microorganism, both genes coding for the same gene product provided that the intron-free gene is capable of expressing the product within a prokaryotic or eukaryotic microorganism. The gene product may be a chemical compound the production of which is desired, for example a protein.

According to a second, preferred aspect of the invention there is provided an intron-free structural gene which encodes PAL or a polypeptide which displays PAL activity. The gene is preferably derived from a PAL - producing strain of a eukaryotic organism, most preferably a strain of R toruloides.

A portion of the genetic DNA polynucleotide sequence of R.toruloides is shown in Fig 3. The methods used by the inventors to determine this sequence are described later. The PAL encoding sequence extends from the location marked "start codon" to the location marked "stop codon", and the introns, six in number are marked IVS 1 to IVS 6. The amino acids encoded by these codons are shown, as also are various restriction sites. The gene of the second aspect of the invention therefore preferably consists of a DNA sequence identical to, related to, derived from or complementary to the sequence of codons from the start codon to the stop codon in Fig 3, from which the six introns IVS 1 to IVS 6 have been deleted, having the following polynucleotide sequence:

ATG GCG CCT CGA CCA ACC TCG CAG TCG CAG GCT CGC ACC TGC CCC ACA
ACC CAG GTC ACG CAG GTC GAC ATC GTC GAG AAG ATG CTC GCC GCG CCG
ACC GAC TCG ACG CTC GAA CTC GAC GGC TAC TCG CTC AAC CTC GGA GAC
GTC GTC TCG GCC GCG AGG AAG GGC AGG CCT GTC CGC GTC AAG GAC AGC
GAC GAG ATC CGC TCA AAG ATT GAC AAA TCG GTC GAG TTC TTG CGC TCG
CAA CTC TCC ATG AGC GTC TAC GGC GTC ACG ACT GGA TTT GGC GGA TCC
GCA GAC ACC CGC ACC GAG GAC GCC ATC TCG CTC CAG AAG GCT CTC CTC
GAG CAC CAG CTC TGC GGT GTT CTC CCT TCG TCG TTC GAC TCG TTC CGC
CTC GGC CGC GGT CTC GAG AAC TCG CTT CCC CTC GAG GTT GTT CGC GGC
GCC ATG ACA ATC CGC GTC AAC AGC TTG ACC CGC GGC CAC TCG GCT GTC
CGC CTC GTC GTC CTC GAG GCG CTC ACC AAC TTC CTC AAC CAC GGC ATC
ACC CCC ATC GTC CCC CTC CGC GGC ACC ATC TCT GCG TCG GGC GAC CTC
TCT CCT CTC TCC TAC ATT GCA GCG GCC ATC AGC GGT CAC CCG GAC AGC
AAG GTG CAC GTC GTC CAC GAG GGC AAG GAG AAG ATC CTG TAC GCC CGC
GAG GCG ATG GCG CTC TTC AAC CTC GAG CCC GTC GTC CTC GGC CCG AAG
GAA GGT CTC GGT CTC GTC AAC GGC ACC GCC GTC TCA GCA TCG ATG GCC
ACC CTC GCT CTG CAC GAC GCA CAC ATG CTC TCG CTC CTC TCG CAG TCG

CTC ACG GCC ATG ACG GTC GAA GCG ATG GTC GGC CAC GCC GGC TCG TTC
CAC CCC TTC CTT CAC GAC GTC ACG CGC CCT CAC CCG ACG CAG ATC GAA
GTC GCG GGA AAC ATC CGC AAG CTC CTC GAG GGA AGC CGC TTT GCT GTC
CAC CAT GAG GAG GAG GTC AAG GTC AAG GAC GAC GAG GGC ATT CTC CGC
CAG GAC CGC TAC CCC TTG CGC ACG TCT CCT CAG TGG CTC GGC CCG CTC
GTC AGC GAC CTC ATT CAC GCC CAC GCC GTC CTC ACC ATC GAG GCC GGC
CAG TCG ACG ACC GAC AAC CCT CTC ATC GAC GTC GAG AAC AAG ACT TCG
CAC CAC GGC GGC AAT TTC CAG GCT GCC GCT GTG GCC AAC ACC ATG GAG
AAG ACT CGC CTC GGG CTC GCC CAG ATC GGC AAG CTC AAC TTC ACG CAG
CTC ACC GAG ATG CTC AAC GCC GGC ATG AAC CGC GGC CTC CCC TCC TGC
CTC GCG GCC GAA GAC CCC TCG CTC TCC TAC CAC TGC AAG GGC CTC GAC
ATC GCC GCT GCG GCG TAC ACC TCG GAG TTG GGA CAC CTC GCC AAC CCT
GTG ACG ACG CAT GTC CAG CCG GCT GAG ATG GCG AAC CAG GCG GTC AAC
TCG CTT GCG CTC ATC TCG GCT CGT CGC ACG ACC GAG TCC AAC GAC GTC
CTT TCT CTC CTC CTC GCC ACC CAC CTC TAC TGC GTT CTC CAA GCC ATC
GAC TTG CGC GCG ATC GAG TTC GAG TTC AAG AAG CAG TTC GGC CCA GCC
ATC GTC TCG CTC ATC GAC CAG CAC TTT GGC TCC GCC ATG ACC GGC TCG
AAC CTG CGC GAC GAG CTC GTC GAG AAG GTG AAC AAG ACG CTC GCC AAG
CGC CTC GAG CAG ACC AAC TCG TAC GAC CTC GTC CCG CGC TGG CAC GAC
GCC TTC TCC TTC GCC GCC GGC ACC GTC GTC GAG GTC CTC TCG TCG ACG
TCG CTC TCG CTC GCC GCC GTC AAC GCC TGG AAG GTC GCC GCC GCC GAG
TCG GCC ATC TCG CTC ACC CGC CAA GTC CGC GAG ACC TTC TGG TCC GCC
GCG TCG ACC TCG TCG CCC GCG CTC TCG TAC CTC TCG CCG CGC ACT CAG
ATC CTC TAC GCC TTC GTC CGC GAG GAG CTT GGC GTC AAG GCC CGC CGC
GGA GAC GTC TTC CTC GGC AAG CAA GAG GTG ACG ATC GGC TCG AAC GTC
TCC AAG ATC TAC GAG GCC ATC AAG TCG GGC AGG ATC AAC AAC GTC CTC
CTC AAG ATG CTC GCT TAG .

It is well known in the field of genetics that DNA sequences which are related to or derived from a defined sequence may encode the same protein or a polypeptide having similar activity to that expressed by the defined sequence. For example the related or derived sequence may lack some bases or may include some additional bases. Also it is known that the genetic code is degenerate, in that several codons may encode the same amino acid. The related or derived sequence may therefore contain some codons which are different to those listed in Fig 3 but which preferably encode the same amino acid. Genes which are related to or derived from this sequence of codons in one or more of these ways are included in the invention.

Genes related to or derived from this sequence may also be defined in terms of the degree of conformity to this sequence. This is preferably as high as possible, ideally 100%, but 70% or higher, eg 85% or higher conformity to that sequence is generally satisfactory.

To enable a gene according to the first or second aspects of the invention to be introduced into a host organism, it is common to include the gene into a recombinant DNA molecule. According to a third aspect of the invention there is therefore provided a recombinant DNA molecule, especially a plasmid, which contains a gene according to the first or second aspects of the invention.

The plasmid according to this aspect of the invention may be used as a vector to introduce the gene into a host and may therefore also contain additional genetic material appropriate to a host into which it is intended to introduce the plasmid. Such genetic material may preferably contain an expression control sequence operatively linked to said gene, and/or transcription/translation signals from other genes appropriate to the organism into which the plasmid is to be introduced and from which expression of the product, eg PAL, is hoped.

The structure of the plasmid according to this aspect of the invention will vary according to the host organism for which it is to be used as a vector, but by positioning the gene of the first or second aspect of the invention downstream of the appropriate regulatory signals,

vectors may be prepared using which expression of R. toruloides PAL may be obtained in any of the currently used production organisms. These include E. coli K12, Bacillus subtilis, Saccharomyces cerevisiae, Pseudomonas putida, Erwinia chrysanthemi and mammalian cell lines.

Similarly the nature of the regulating DNA sequences immediately upstream of the PAL cDNA coding region in the plasmid will be composed of appropriate, characterised transcription/translation signals. For example for use in S. cerevisae a ribosome binding site (conforming to the sequence CCACCTT) may be positioned at the appropriate position upstream of the translational start of the PAL gene, and powerful transcriptional signals, such as those derived from the S. cerevisae phosphoglycerate kinase and mating facter genes, placed 5' to the ribosome binding site. The plasmid itself may use standard replicons (eg 2p) and selectable markers (e.g. Leu2, Trp etc). Similarly, for use in E. coli use will be made of the PL, tac trp, rac or lac promoters, with appropriate bacterial ribosome binding sites, and plasmids based on ColE1 (e.g. pBR322 and pUC plasmids), RSF1010, and runaway replicons of RI. As the introns present in the natural PAL gene act as a barrier to the expression of PAL in organisms other than R. toruloides, the invention may be used to produce PAL in a wide range of procaryotic and eukaryotic hosts which are unable to express the natural PAL gene due to the presence of the 6 introns.

In accordance with a fourth aspect of the invention there is provided a host organism, especially a strain of E. coli, Erwinia sp., Clostridia sp., Streptomyces sp., B. subtilis, B. stearothermophilus, Pseudomonas, other microorganisms such as bacilli, yeasts, other fungi, animal or plant hosts, and preferably a prokaryotic host, transformed with at least one recombinant DNA molecule according to the third aspect.

The invention also provides a process for the preparation of a gene from which introns have been deleted which includes the steps of:

(i) isolating PAL mRNA from a strain of R. toruloides,

(ii) synthesising two intron-free complementary DNA ('cDNA') sequences from the mRNA, the two cDNA sequences each containing a portion of a gene which encodes PAL or a polypeptide

which displays PAL activity, the two portions together containing the 3' and the 5' ends of the gene.

(iii) joining the two cDNA sequences proposed in (ii) to form an intron-free structural gene which encodes PAL or a polypeptide which displans PAL activity.

The method used in step (ii) may use a cloning method which forms the cDNA sequences contained in plasmids. In such a case the sequences may be isolated from the plasmids which contain them by cleavage of the plasmids at a suitable restriction site, followed by ligation of the two sequences to form the gene. The gene may then be combined with other genetic material to form a plasmid containing it for example following cleavage of a suitable known plasmid such as pUC9 at appropriate sites. If desired the gene may then in turn be excised from this plasmid and combined with yet other genetic material to form other plasmids which may be used as vectors. Standard recombinant DNA techniques, familiar to those skilled in the art may be used for the process of the invention.

The gene and/or plasmid produced in step (iii) of this process is preferably one of the genes or plasmids encompassed by the second and/or third aspects of the invention, and the cDNA sequences produced in step (ii) are consequently preferably portions of these. The cDNA sequences produced in step (ii) and plasmids containing them are further aspects of the invention.

The invention therefore also includes DNA polynucleotide sequences, eg plasmids, the same as or substantially the same as or derived from or related to those produced by the process of the invention.

The invention will now be described by way of examole only with reference to the accompanying figures:

Fig 1 is a schematic diagram illustrating how the genetic DNA carrying the PAL gene was sequenced.

Fig 2 illustrates the production of the two plasmids carrying the PAL gene which lack the intron sequences of the natural gene.

Fig 3 shows the complete nucleotide sequence of the genomic clone, the intron sequences removed in the invention being labelled IVS1 to IVS6, and the corresponding amino acid sequence of PAL.

Fig 4 shows the formation of the recombinant plasmid pPAL 3 containing the intron-free gene, by combination of the two cDNA plasmids pPAL1 and pPAL2.

Figs 5 & 6 illustrates the DNA nucleotide sequences of the overlapping cDNA clones pPAL1 and pPAL 2 respectively.

Fig 7 shows the expression of PAL protein from the plasmid pPAL4.

In this description and the figures the following abbreviations are used:

| Amino acid | symbol | Nucleotide bases | symbol |
|---|---|---|---|
| Alanine | Ala | Uracil | U |
| Arginine | Arg | Thymine | T |
| Asparagine | Asn | Cytosine | C |
| Aspartic acid | Asp | Adenine | A |
| Asn + Asp | Asx | Guanine | G |
| Cysteine | Cys | | |
| Glutamine | Gln | | |
| Glutamic Acid | Glu | | |
| Gln + Glu | Glx | | |
| Glycine | Gly | | |
| Histidine | His | | |
| Isoleucine | Ile | | |
| Leucine | Leu | | |
| Lysine | Lys | | |
| Methionine | Met | | |
| Phenylalanine | Phe | | |
| Proline | Pro | | |
| Serine | Ser | | |
| Threonine | Thr | | |
| Tryptophan | Trp | | |

| Amino acid | Symbol |
|---|---|
| Tyrosine | Tyr |
| Valine | Val |

Referring to Figs 1 to 6 in more detail:

In Fig 1. Region 2 was isolated from the appropriate clone (pHG3), circularised by treatment with T4 DNA ligase, fragmented by sonication, and fragments of between 500 and 1000 bp inserted into M13mp8. Region (1) was inserted into M13mp8 and M13mp9 as various specific fragments utilising the restriction sites BamHI, BclI and SalI. The sequence of the DNA spanning the BamHI site was obtained by cloning the indicated fragment (3) into M13mp8.

In Fig 2. Clone 1 (pPAL1) was obtained by the method of Heidecker and Messing (1983). Total mRNA from pal-induced R. toruloides cells was annealed to oligo(dT)-tailed pUC9, and the first strand cDNA copy synthesised using reverse transcriptase in the presence of all four dNTP's. The newly synthesised strands were tailed with oligo(dC) using terminal deoxynucleotidyl tranferase. Following fractionation by an alkaline sucrose gradient, single-stranded plasmid DNA carrying cDNA sequences were annealed to denatured oligo(dG)-tailed pUC9 and the second strand synthesised using DNA polymerase (Klenow) and the addition of all four dNTP's. Clone 2 (pPAL2) was constructed using the procedure of Gubler and Hoffman (1982). The first strand cDNA copy was synthesised using reverse transcriptase and a 19-mer oligodeoxynucleotide primer (GATCAGAGGGTTGTCGGTC) complementary to pal mRNA. The RNa within the RNA-DNa hybrid was then nicked with RNase H and the RNA strand replaced with DNA by E. coli DNA polymerase, utilsing the nicked RNA as a primer. The double stranded DNA was then blunt ended by the action of T4 DNA polymerase, tailed with oligo(dC), and annealed to oligo(dG) tailed pBR322. cDNA clones produced using both methods were transformed into E. coli JM83, and colonies screened for Pal cDNa sequences using [ $\alpha$-$^{32}$p] dATP-labelled pHG3 restriction fragments.

In Fig 3. The determined amino acid sequences of the 5 randomly derived peptide fragments are indicated by overlining of the relevant residues. The introns are labelled IVS 1-6, and the sequence common to all 6 is indicated by underlining of the relevant region. A dashed overline in the 5' non-coding region represents the TC rich region of the sequence, while the under- and overlining immediately downstream marks a repetitive region. The sequence extends from the most leftward BclI site of Figure 1, to the 3' end of cDNA clone PPAL1 (Figure 2).

In Fig 4. The single line of the pPAL1 and pPAL3 circular maps represent pUC9 and pUC8 derived DNA, respectively, while the single line of pPAL2 represents pBR322 DNA (see Fig. 2 for construction of pPAL1 and pPAL2). The double line of pPAL2 represents the "intron-free" 5' end of the PAL gene, while the thick line of pPAL1 represents the 3' end of the gene. The 5' end of the PAL gene was isolated as a 1.0kb PstI - FspI fragment from pPAL2 and ligated to a 1.25kb FspI - BamHI fragment, isolated from pPAL1, which carried the 3'end of the gene. The ligated fragment was inserted between the BamHI and PstI sites of pUC8 to yield pPAL3. The positions of the PAL gene translational start (ATG) and stop (TAG) codon (see Fig. 3) are marked by arrows. The orientation of insertion of the PAL gene is such that transcriptional read through from the vector borne lac promoter (lac po) will not occur.

In Fig 5 & 6. The Fsp 1 site used to join these two clones to form pPAL 3 is indicated by underlining of the relevant nucleotides.

In Fig 7. The plasmid pPAL4 contains the complete PAL gene from pPAL 3 (Fig 4) cloned into pUC9 as an EcoRI - Hind III fragment such that transcriptional read through from the adjacent lac promoter can occur. Gene product formation was assessed using a plasmid-directed in vitro translation kit obtained from Amersham International PLC. Samples in the numbered tracks are as follows; 1, no DNA added; 2, plasmid pUC9; 3, plasmid pPAL4. Molecular weights of the protein markers are given as $M_r$.

In the following description reference will be made to the following general procedures:

MICROBIAL STRAINS AND PLASMIDS

Microbial strains and plasmids used in accordance with the invention are listed in Table 1.

MEDIA

E. coli strains were cultured in L-broth (1% tryptone, 0.5% yeast extract, 0.5% NaCl). Media were solidified with the addition of 2% (w/v) Bacto-agar (Difco). Ampicillin (100 $\mu$g $ml^{-1}$ was used for the selection and growth of transformants. Functional $\beta$ -galactosidase was detected by the addition of 5-bromo-4-chloro-indoyl-$\beta$ -D-galactoside (X-Gal) to a final concentration of 2 $\mu$g $ml^{-1}$

CHEMICALS

[$\alpha^{32}$P] dATP and the cDNA synthesis kit were obtained from Amersham International. Agarose, restriction enzymes, T4 DNA ligase, terminal deoxynucleotidyl transferase and 17mer universal sequence primer were purchased from Bethesda Research Laboratories. Klenow DNA polymerase was from Boehringer Mannheim, while dT tailed pUC9 was from PL-Biochemicals. Reverse transcriptase was purchased from Anglicon Biotechnology Ltd. while all other reagents were obtained from Sigma Chemical Co. or BDH.

**TABLE 1 Microbial Strains and Vectors Used**

| | | |
|---|---|---|
| Strain | | |
| *E. coli* JM83 | *ara*, (*lac-pro*) *rpsL*, *thi*, 080d *lacI* ZM15 | Vieria and Messing (1982) |
| *E. coli* JM101 | (*lac-pro*), *supE*, *thi*/ F*lacI* ZM15 *traD pro* | Messing and Vieria (1982) |
| Plasmids | | |
| pUC9 | $Amp^R$ | Vieria and Messing (1982) |
| pBR322 | $Amp^R$ $Tet^R$ | Bolivar *et al.* (1977) |
| pGH3 | $Amp^R$ (PAL genomic clone) | Gilbert *et al.* (1985) |
| pPAL1 | $Amp^R$ (3' end PAL cDNA clone) | Novel plasmids |
| pPAL2 | $Amp^R$ (5' end PAL cDNA clone) | Novel plasmids |
| pPAL3 | $Amp^R$ (entire PAL cDNA gene) | Novel plasmids |
| pPAL4 | $Amp^R$ (entire PAL cDNA gene) | Novel plasmids |
| Bacteriophage | | |
| M13mp8 | | Messing and Vieria (1982) |
| M13mp9 | | Messing and Vieria (1982) |

DNA MANIPULATIONS

All restriction enzymes and DNA/RNA modifying enzymes were used in the buffers and under the conditions recommended by the suppliers. Plasmid transformation techniques and all manipulation of DNA have previously been described (Minton et al., 1984).

PLASMID DNA ISOLATION

E. coli plasmids were purified from 1 litre of L-broth cultures containing ampicillin by "Brij lysis" and subsequent CsCl density gradient centrifugation (Clewell and Helinski, 1969). The rapid boiling method of Holmes and Quigley (1981) was employed for small scale plasmid isolation screening purposes.

TEMPLATE GENERATION BY SONICATION

The DNA to be sequenced was fragmented into random blunt-ended fragments by the procedure of Deininger (1983). The fragments obtained were cloned into the SmaI site of M13mp8 and template DNA prepared as described by Sanger et al (1980).

NUCLEOTIDE SEQUENCING

Nucleotide sequencing was undertaking by the dideoxy method of Sanger et al (1980). The data obtained was compiled into a complete sequence using the computer programmes of Staden ( 1980 ).

ISOLATION OF PAL mRNA

PAL mRNA was isolated as has previously been described (Gilbert et al., 1985) employing publicly available R. toruloides strain IFO 0559 (equivalent to NCYC 1589 deposited at the National Collection of Yeast Cultures, Norwich (GB) under the terms of the Budapest Treaty on 8 September 1986).

CDNA CLONING (SYNTHESIS)

i) Heidecker-Messing Method

The method utilised was essentially as described by Heidecker and Messing (1983). Total mRNA from PAL induced R. toruloides cells was annealed to oligo dT tailed pUC9 and the first strand cDNA copy synthesised using reverse transcriptase in the presence of all 4 dNTP's. The newly synthesised strands were tailed with oligo dC using terminal deoxynucleotidyl transferase. Following fractionation by an alkaline sucrose gradient, single stranded plasmid DNA carrying cDNA sequences were annealed to denatured oligo dG tailed pUC9 and the second strand synthesised using Klenow DNA polymerase and the addition of all 4 dNTP's.

ii) Gubler-Hoffman Method

The second method employed in the synthesis of cDNA was that of Gubler and Hoffman (1983). The first strand cDNA copy was synthesised using reverse transcriptase and a 19-mer primer (GATGAGAGGGTTGTCGGTC) complementary to PAL mRNA. The RNA within the RNA-DNA hybrid was then "nicked" with RNaseH and the RNA strand replaced with DNA by E. coli DNA polymerase, utilising the nicked RNA as a primer. The double stranded DNA was then blunt-ended by the action of T4 DNA polymerase, tailed with oligo dC and annealed to oligo-dG tailed pBR322.

DETECTION OF PAL cDNA CLONES

Plasmid DNA carrying cDNA inserts were transformed into E. coli JM83 and the $Amp^R$ transformants screened for PAL specific DNA. This was undertaken by in situ colony hyridisation (Grunstein and Hogness, 1975), utilising radio labelled pHG3 DNA subfragments carrying portsion of the PAL gene.

AMINO ACID SEQUENCING

Peptide fragments of purified (according to Gilbert et al., (1985) PAL protein were isolated as previously described (Minton et al., 1984). Amino acid sequencing was undertaken using automated Edman degradation

using an Applied Biosystems gas phase sequencer, model 470A.

IN VITRO TRANSLATION

The bacterial ion-free coupled transcription-translation system used was a modification of that first described by De Vries and Zubay (1967). The E. coli S-30 extract and the supplement solutions required for in vitro expression of genes contained on a bacterial plasmid were purchased as a kit from Amersham International PLC. Proteins produced were labelled with $^{35}$S-methionine (Amersham), and analysed by SDS-PAGE on 12% acrylamide gels (Laemmli, 1970). Gels were dried prior to autoradiography for 16 hours.

1. NUCLEOTIDE SEQUENCING OF THE PAL GENOMIC CLONE

The PAL gene was previously shown (Gilbert et al., 1985) to occupy a 2.5 kb region of DNA within a 6.7 kb BclI fragment cloned into pUC8 to yield the recombinant plasmid pHG3 (see Fig. 1). The majority of the gene resided on a 3 kb BamHI fragment, while the remaining 5' end of the gene lay on a 0.7 kb BamHI-BclI fragment. Accordingly, the 3 kb fragment was isolate from an appropriate clone (fragment 2, Fig. 1) and random subfragments, generated by sonication (Deininger, 1983), cloned into M13mp8. A total of some 250 templates were prepared and sequenced, the data obtained being compiled into a complete sequence using the computer programmes of Staden (1980). The sequence of the 5' end of the gene was obtained by the site directed cloning of the relevant BclI-BamHI, BclI-SalI and BamHI-SalI fragments (region 1, Fig. 1.) into the appropriate sites of M13mp8 and M13mp9. Sequence determination of the DNA spanning the BamHI site was achieved by cloning the SalI-XhoI fragment (3) indicated in Fig. 1.

The translation of the appropriate DNA strand of the sequenced region indicated that an open reading frame (ORF) capable of coding for PAL was not present. Confirmation that this region does encode PAL, however, was obtained by comparing the translated amino acid sequences with the determined sequence of 5 randomly derived peptide fragments. All 5 peptide sequences were located within the translated sequence but occurred in various translational reading frames (Fig. 2.). The absence of a contiguous ORF suggested that, in common with other fungal genes, the PAL gene contains introns.

2. ISOLATION OF cDNA CLONES CARRYING THE PAL GENE

To enable the identification of the PAL intervening sequences we elected to reclone the gene from cDNA. In the initial experiments the procedure of Heidecker-Messing (1983) was adopted, utilising the vector pUC9 and purified PAL mRNA. Clones carrying PAL DNA sequences were identified utilising the 3 kb BamHI-BclI fragment of pHG3 as a DNA probe. The largest clone obtained, pPAL1, contained some 1.3 kb from the 3' end of the PAL gene (Fig. 3). The 5' end of the gene was obtained by cloning C-tailed cDNA, prepared by the method Gubler and Hoffman (1983), into

G-tailed pBR322, to yield pPAL2. In this case the primer utilised during first strand synthesis was a synthesised 19-mer oligonucleotide complementary to the PAL coding strand 150 bp downstream from the 5' end of the previously obtained cDNA (see Fig. 3.). The nucleotide sequence of the two cDNA clones was determined by site directed cloning of appropriate restriction fragments into M13mp8 and M13mp9.

3. **IDENTIFICATION OF THE PAL INTRONS**

Sequence determination of the 2 clones confirmed the presence of 6 introns within the PAL coding sequence. Thus the 6 regions of DNA labelled IVS1 to IVS6 were completely absent from the appropriate regions of pPAL1 and pPAL2. Examination of the 6 missing regions revealed that they all contained the nucleotides CAG at their 3' ends, exhibiting perfect agreement to the consensus intron accepter sequence generally observed in eukaryotic genes (Mount, 1982). A number of the sequences at the 5' end of some of these introns demonstrated less conformity to the eukaryotic consensus donor sequence (GTA/GAGT). Thus the donor sequences of IVS 2, 4 and 5 were GTGCGT, GTGCGC and GTGCGC respectively. The introns of eukaryotic genes have been generally shown to contain sequences necessary for the accurate splicing of the intervening non-coding regions. Seqeunces conforming to consensus sequences observed in the introns of other eukaryotics (e.g. TACTTAACA in S. cerevisae; see Orbach et al., 1986) are not present in the R. toruloides introns. In their place a sequence is present conforming to the consensus G/ANG/CTGAC (the relevant sequence within each intron has been underlined in Fig. 3). Such a sequence may be specific to R. toruloides and closely related organisms.

The PAL gene has been shown not to express in either E. coli (Gilbert et al., 1985) or S. cerevisae (Tully and Gilbert, 1985). The reason for lack of expression in the former is undoubtedly due to the presence of introns in the PAL gene. Furthermore, although S. cerevisae is capable of splicing introns, the differences in the nucleotide sequences of the PAL introns and those found in S. cerevisae intron probably explains the inability of this yeast to express the R. toruloides PAL gene.

## 4. DERIVATION OF A CONTIGUOUS cDNA PAL GENE

The procedure utilised in the cloning of the PAL gene from cDNA had resulted in two clones, pPAL1, which carried the 3' end of the gene, and pPAL2, carrying the 5' end of the gene. A third plasmid was constructed, carrying the entire PAL structural gene by amalgamating the inserts of the above two plasmids. This was achieved by isolating a 1.0kb FspI - PstI fragment from pPAL2, carrying the 5' end of PAL, and ligating it to a 1.25kb FspI - BamHI fragment carrying the 3' end of the gene isolated from pPAL1. The ligated DNA was then inserted into pUC9 cleaved with PstI and BamHI (Fig. 4). The plasmid pPAL3 therefore carries the entire PAL structural gene, but lacks all 6 introns found in the natural R. toruloides chromosomal gene.

## 5. SYNTHESIS OF PAL PROTEIN

The fragment containing the complete intron-free PAL gene from pPAL3 has been cloned into the pUC plasmids in both orientations relative to the lac promoter, to give pPAL3 (pUC8) and pPAL4 (pUC9). In pPAL4, the PAL gene is in phase with the lacZ promoter from pUC9, and synthesis of PAL protein has been demonstrated in a plasmid-directed in vitro translation system. This is shown in Fig.7. With the PAL gene in the opposite orientation (pPAL3) no PAL protein is produced. We are currently developing vector systems to enable us to express the PAL gene in Saccharomyces cerevisiae.

Ambrus, C.M., J.L. Ambrus, C. Hoarvath, H. Pedersen, S. Sharma, C. Kent, E. Mirand, R. Guthrie, and T. Paul. 1978. Phenylalanine depletion for the management of phenylketonuria: use of enzyme reactors with immobilized enzymes. Science 201: 837-839.

Attridge, T.H., C.B. Jonhson, and H. Smith. 1974. Density-labelling evidence for the phytochrome-mediated activation of phenylalanine ammonia-lyase in mustard cotyledons. Biochim. Biophys. Acta. 343: 440-451.

Boulivar, F., R.L. Rodriquez, P.J. Green, M.C. Betlach, H.L. Heyneker and H.W. Boyer. 1977. Construction and characterisation of new cloning vehicles II. A multipurpose cloning system. Gene 2: 95-113.

Clewell, D.B., and D.R. Helinski. 1969. Supercoiled circular DNA-protein complex in E. coli. Purification and induced conversion to an open circular form. Proc. Natl. Acad. Sci. USA 62: 1159-1166.

DeVries, J.K., and G. Zubay. 1967. DNA-directed peptide synthesis II. The synthesis of the -fragment of the enzyme β -galactosidase. Proc. Natl. Acad. Sci. USA 57: 1010-1012.

Gilbert, H.J., J.R. Stephenson, and M. Tully. 1983. Control of synthesis of functional mRNA coding for phenylalanine ammonia-lyase for Rhodosporidium toruloides. J. Bacteriol. 153: 1147-1154.

Gilbert, H.J., and M. Tully. 1982. Synthesis and degradation of phenylalanine ammonia-lyase of Rhodosporidium toruloides. J. Bacteriol. 150: 498-505.

Gilbert, H.J., I.N. Clarke, R.K. Gibson, J.R. Stephenson, and M. Tully. 1985. Molecular cloning of the phenylalanine ammonia lyase gene from Rhodosporidium toruloides in Escherichia coli K-12. J. Bacteriol. 161: 314-320.

Grunstein, M., and D.S. Hogness. 1975. Colony hybridisation: a method for the isolation of cloned DNA's that contain a specific gene. Proc. Natl. Acad. Sci. USA 72: 3961-3965.

Gubler, U., and B.J. Hoffman. 1983. A simple and very efficient method of generating cDNA libraries. Gene 25: 263-269.

Heidecker, G., and J. Messing. 1983. Sequence analysis of zein cDNA; obtained by an efficient mRNA cloning method. nacl. Acids. Res. 11: 4891-4906.

Holmes, D.S., and M. Quigley. 1981. A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. 114: 193-197.

Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the lead of bacteriophage T4. Nature 227: 680-685.

Marusich, W.C., R.A. jensen, and L.O. Zamir. 1981. Induction of

L-phenylalanine ammonia-lyase during utilization of phenylalanine as carbon or nitrogen source in Rhodotorula glutinis. J. Bacteriol. 146: 1013-1019.

Messing, J., and J. Vieria. 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. 19: 269-276.

Minton, N.P., T. Atkinson, C.J. Bruton and R.F. Sherwood. 1984. The complete nucleotide sequence of the Pseudomonas gene coding for carboxypeptidase $G_2$. Gene 31: 31-38.

Mount, S.M. 1982. A catalogue of splice junction sequences. Nacl. Acids. Res. 10: 459-472.

Orbach, M.J., E.B. Porro, and C. Yanotsky. 1986. Cloning and characterisation of the gene for -tublin from a benomyl-resistant mutant of Neurospora crassa and its use as a dominant selectable marker. Mol. Cell. Biol. 6: 2452-2461.

Sanger, F., A.R. Coulson, B.G. Barrell, A.J.H. Smith and B.A. Roe. 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. 143: 161-178.

Schroder, J., F. Kreuzaler, E. Schafer, and K. Hahlbrock. 1979. Concomitant induction of phenylalanine ammonia-lyase and flavanone synthase in irradiated plant cells. J. Biol. Chem. 254: 57-65.

Tan, S.C. 1980. Phenylalanine ammonia-lyase and the phenylalanine ammonia-lyase inactivating system: effects of light, temperature, and mineral deficiencies. Aust. J. Plant. Physiol. 1: 159-167.

Tully, M., and H.J. Gilbert. 1985. Transformation of Rhodosporidium toruloides. Gene 36: 235-240.

Vieria, J., and J. Messing. 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19: 259-260.

Yamada, S., K. Kabe, N. Izuo, K. Nakamichi, and I. Chibata. 1981. Production of L-phenylalanine from trans-cinnamic acid with Rhosdotorula glutinis containing L-phenylalanine ammonia-lyase activity. Appl. Environ. Microbiol. 42: 773-778.

CLAIMS

1. A gene, characterised in that it is an intron-free structural gene, derived from a corresponding intron-containing structural gene from a eukaryotic microorganism, both genes encoding the same product provided that the intron-free gene is capable of expressing the product within a prokaryotic or eukaryotic microorganism.

2. A gene as claimed in claim 1, characterised in that the gene encodes phenylalanine ammonia lyase ('PAL') or a polypeptide which displays PAL activity.

3. A gene as claimed in claim 2, characterised in that it is derived from an intron-containing gene of a PAL-producing strain of a eukaryotic microorganism.

4. A gene as claimed in claim 3, characterised in that the micro-organism is R toruloides.

5. A gene characterised in that it has a structure identical to, related to, derived from or complementary to the following polynucleotide sequence:

ATG GCG CCT CGA CCA ACC TCG CAG TCG CAG GCT CGC ACC TGC CCC ACA
ACC CAG GTC ACG CAG GTC GAC ATC GTC GAG AAG ATG CTC GCC GCG CCG
ACC GAC TCG ACG CTC GAA CTC GAC GGC TAC TCG CTC AAC CTC GGA GAC
GTC GTC TCG GCC GCG AGG AAG GGC AGG CCT GTC CGC GTC AAG GAC AGC
GAC GAG ATC CGC TCA AAG ATT GAC AAA TCG GTC GAG TTC TTG CGC TCG
CAA CTC TCC ATG AGC GTC TAC GGC GTC ACG ACT GGA TTT GGC GGA TCC
GCA GAC ACC CGC ACC GAG GAC GCC ATC TCG CTC CAG AAG GCT CTC CTC
GAG CAC CAG CTC TGC GGT GTT CTC CCT TCG TCG TTC GAC TCG TTC CGC
CTC GGC CGC GGT CTC GAG AAC TCG GTT CCC CTC GAG GTT GTT CGC GGC
GCC ATG ACA ATC CGC GTC AAC AGC TTG ACC CGC GGC CAC TCG GCT GTC
CGC CTC GTC GTC CTC GAG GCG CTC ACC AAC TTC CTC AAC CAC GGC ATC
ACC CCC ATC GTC CCC CTC CGC GGC ACC ATC TCT GCG TCG GGC GAC CTC
TCT CCT CTC TCC TAC ATT GCA GCG GCC ATC AGC GGT CAC CCG GAC AGC
AAG GTG CAC GTC GTC CAC GAG GGC AAG CAG AAG ATC CTG TAC GCC CGC
GAG GCG ATG GCG CTC TTC AAC CTC GAG CCC GTC GTC CTC GGC CCG AAG
GAA GGT CTC GGT CTC GTC AAC GGC ACC GCC GTC TCA GCA TCG ATG GCC
ACC CTC GCT CTG CAC GAC GCA CAC ATG CTC TCG CTC CTC TCG CAG TCG

5. (contd.)

CTC ACG GCC ATG ACG GTC GAA GCG ATG GTC GGC CAC GCC GGC TCG TTC
CAC CCC TTC CTT CAC GAC GTC ACG CGC CCT CAC CCG ACG CAG ATC GAA
GTC GCG GGA AAC ATC CGC AAG CTC CTC GAG GGA AGC CGC TTT GCT GTC
CAC CAT GAG GAG CAG GTC AAG GTC AAG GAC GAC GAG GGC ATT CTC CGC
CAG GAC CGC TAC CCC TTG CGC ACG TCT CCT CAG TGG CTC GGC CCG CTC
GTC AGC GAC CTC ATT CAC GCC CAC GGC GTC CTC ACC ATC GAG GCC GGC
CAG TCG ACG ACC GAC AAC CCT CTC ATC GAC GTC GAG AAC AAG ACT TCG
CAC CAC GGC GGC AAT TTC CAG GCT GCC GCT GTG GCC AAC ACC ATG GAG
AAG ACT CGC CTC GGG CTC GCC CAG ATC GGC AAG CTC AAC TTC ACG CAG
CTC ACC GAG ATG CTC AAC GCC GGC ATG AAC CGC GGC CTC CCC TCC TGC
CTC GCG GCC GAA GAC CCC TCG CTC TCC TAC CAC TGC AAG GGC CTC GAC
ATC GCC GCT GCG GCG TAC ACC TCG GAG TTG GGA CAC CTC GCC AAC CCT
GTG ACG ACG CAT GTC CAG CCG GCT GAG ATG GCG AAC CAG GCG GTC AAC
TCG CTT GCG CTC ATC TCG GCT CGT CGC ACG ACC GAG TCC AAC GAC GTC
CTT TCT CTC CTC CTC GCC ACC CAC CTC TAC TGC GTT CTC CAA GCC ATC
GAC TTG CGC GCG ATC GAG TTC GAG TTC AAG AAG CAG TTC GGC CCA GCC
ATC GTC TCG CTC ATC GAC CAG CAC TTT GGC TCC GCC ATG ACC GGC TCG
AAC CTG CGC GAC GAG CTC GTC GAG AAG GTG AAC AAG ACG CTC GCC AAG
CGC CTC GAG CAG ACC AAC TCG TAC GAC CTC GTC CCG CGC TGG CAC GAC
GCC TTC TCC TTC GCC GCC GGC ACC GTC GTC GAG GTC CTC TCG TCG ACG
TCG CTC TCG CTC GCC GCC GTC AAC GCC TGG AAG GTC GCC GCC GCC GAG
TCG GCC ATC TCG CTC ACC CGC CAA GTC CGC GAG ACC TTC TGG TCC GCC
GCG TCG ACC TCG TCG CCC GCG CTC TCG TAC CTC TCG CCG CGC ACT CAG
ATC CTC TAC GCC TTC GTC CGC GAG GAG CTT GGC GTC AAG GCC CGC CGC
GGA GAC GTC TTC CTC GGC AAG CAA GAG GTG ACG ATC GGC TCG AAC GTC
TCC AAG ATC TAC GAG GCC ATC AAG TCG GGC AGG ATC AAC AAC GTC CTC
CTC AAG ATG CTC GCT TAG

which encodes PAL or a polypeptide which displays PAL activity.

6. A gene as claimed in claim 5 characterised in that it lacks some bases or includes some additional bases or has some of the listed codons replaced by other codons, provided that the gene encodes PAL or a polypeptide displaying PAL activity.

7. A recombinant DNA molecule characterised in that it contains a gene as claimed in any one of claims 1 to 4.

8. A recombinant DNA molecule characterised in that it contains a gene as claimed in claim 5 or claim 6.

9. A molecule as claimed in claim 8 characterised in that it is a plasmid.

10. A plasmid as claimed in claim 9 characterised in that it is a vector and also contains an expression control sequence operatively linked to the gene, and/or transcription/translation signals appropriate of PAL, or a polypeptide which displays PAL activity, from E. coli K12, Bacillus subtilis, Saccharomyces cerevisae, Preudomonas putida, Erwinia chrysanthemi or mammalian cell lines.

11. A molecule as claimed in claim 10 characterised in that is contains a ribosome binding site upstream of the translational start of the gene and transcriptional signals derived from the S cerevisae phosphoglycerate kinase and mating facter genes placed 5' to the ribosome binding site.

12. A recombinant DNA molecule characterised in that it consists of a gene as claimed in claim 5 inserted into the plasmid pUC9 with the gene in phase with the lac Z promoter of pUC 9.

13. A host microorganism characterised in that it is transformed with a recombinant DNA molecule as claimed in claim 8.

14. A process for the preparation of a gene from which introns have been deleted characterised in that it includes the steps of:

(i) isolating PAL mRNA from a strain of R. toruloides.

(ii) synthesising two intron-free cDNA sequences from the mRNA, the two cDNA sequences each containing a portion of a gene which encodes pAL or a polypeptide which displays PAL activity, the two portions together containing the 3' and the 5' ends of the gene,

(iii) joining the two cDNA sequences to form an intron-free structural gene which encodes PAL or a polypeptide which displays PAL activity.

15. A polynucleotide sequence characterised in that it is a portion of an intron-free gene which encodes PAL or a polypeptide which displays PAL activity and contains the 3' or the 5' end of the gene.

16. A sequence as claimed in claim 15 characterised in that it contains a polynucleotide sequence identical to, related to or derived from the following polynucleotide sequence:

ATG GCG CCT CGA CCA ACC TCG CAG TCG CAG GCT CGC ACC TGC CCC ACA
ACC CAG GTC ACG CAG GTC GAC ATC GTC GAG AAG ATG CTC GCC GCG CCG
ACC GAC TCG ACG CTC GAA CTC GAC GGC TAC TCG CTC AAC CTC GGA GAC
GTC GTC TCG GCC GCG AGG AAG GGC AGG CCT GTC CGC GTC AAG GAC AGC
GAC GAG ATC CGC TCA AAG ATT GAC AAA TCG GTC GAG TTC TTG CGC TCG
CAA CTC TCC ATG AGC GTC TAC GGC GTC ACG ACT GGA TTT GGC GGA TCC
GCA GAC ACC CGC ACC GAG GAC GCC ATC TCG CTC CAG AAG GCT CTC CTC
GAG CAC CAG CTC TGC GGT GTT CTC CCT TCG TCG TTC GAC TCG TTC CGC
CTC GGC CGC GGT CTC GAG AAC TCG CTT CCC CTC GAG GTT GTT CGC GGC
GCC ATG ACA ATC CGC GTC AAC AGC TTG ACC CGC GGC CAC TCG GCT GTC
CGC CTC GTC GTC CTC GAG GCG CTC ACC AAC TTC CTC AAC CAC GGC ATC
ACC CCC ATC GTC CCC CTC CGC GGC ACC ATC TCT GCG TCG GGC GAC CTC
TCT CCT CTC TCC TAC ATT GCA GCG GCC ATC AGC GGT CAC CCG GAC AGC
AAG GTG CAC GTC GTC CAC GAG GGC AAG GAG AAG ATC CTG TAC GCC CGC
GAG GCG ATG GCG CTC TTC AAC CTC GAG CCC GTC GTC CTC GGC CCG AAG
GAA GGT CTC GGT CTC GTC AAC GGC ACC GCC GTC TCA GCA TCG ATG GCC
ACC CTC GCT CTG CAC GAC GCA CAC ATG CTC TCG CTC CTC TCG CAG TCG
CTC ACG GCC ATG ACG GTC GAA GCG ATG GTC GGC CAC GCC GGC TCG TTC
CAC CCC TTC CTT CAC GAC GTC ACG CGC CCT CAC CCG ACG CAG ATC GAA
GTC GCG GGA AAC ATC CGC AAG CTC CTC GAG GGA AGC CGC TTT GCT GTC
CAC CAT GAG GAG GAG GTC AAG GTC AAG GAC GAC GAG GGC ATT CTC CGC
CAG GAC CGC TAC CCC TTG CGC ACG .

17. A sequence as claimed in claim 15 characterised in that it contains a polynucleotide sequence identical to, related to or derived from the following polynucleotide sequence:

17. (contd.)

TCT CCT CAG TGG CTC GGC CCG CTC
GTC AGC GAC CTC ATT CAC GCC CAC GCC GTC CTC ACC ATC GAG GCC GCC
CAG TCG ACG ACC GAC AAC CCT CTC ATC GAC GTC GAG AAC AAG ACT TCG
CAC CAC GGC GGC AAT TTC CAG GCT GCC GCT GTG GCC AAC ACC ATG GAG
AAG ACT CGC CTC GGG CTC GCC CAG ATC GGC AAG CTC AAC TTC ACG CAG
CTC ACC GAG ATG CTC AAC GCC GGC ATG AAC CGC GGC CTC CCC TCC TGC
CTC GCG GCC GAA GAC CCC TCG CTC TCC TAC CAC TGC AAG GGC CTC GAC
ATC GCC GCT GCG GCG TAC ACC TCG GAG TTG GCA CAC CTC GCC AAC CCT
GTG ACG ACG CAT GTC CAG CCG GCT GAG ATG GCG AAC CAG GCG GTC AAC
TCG CTT GCG CTC ATC TCG GCT CGT CGC ACG ACC GAG TCC AAC GAC GTC
CTT TCT CTC CTC CTC GCC ACC CAC CTC TAC TGC GTT CTC CAA GCC ATC
GAC TTG CGC GCG ATC GAG TTC GAG TTC AAG AAG CAG TTC GGC CCA GCC
ATC GTC TCG CTC ATC GAC CAG CAC TTT GGC TCC GCC ATG ACC GGC TCG
AAC CTG CGC GAC GAG CTC GTC GAG AAG GTG AAC AAG ACG CTC GCC AAG
CGC CTC GAG CAG ACC AAC TCG TAC GAC CTC GTC CCG CGC TGG CAC GAC
GCC TTC TCC TTC GCC GCC GGC ACC GTC GTC GAG GTC CTC TCG TCG ACG
TCG CTC TCG CTC GCC GCC GTC AAC GCC TGG AAG GTC GCC GCC GCC GAG
TCG GCC ATC TCG CTC ACC CGC CAA GTC CGC GAG ACC TTC TGG TCC GCC
GCG TCG ACC TCG TCG CCC GCG CTC TCG TAC CTC TCG CCG CGC ACT CAG
ATC CTC TAC GCC TTC GTC CGC GAG GAG CTT GGC GTC AAG GCC CGC CGC
GGA GAC GTC TTC CTC GGC AAG CAA GAG GTG ACG ATC GGC TCG AAC GTC
TCC AAG ATC TAC GAG GCC ATC AAG TCG GGC AGG ATC AAC AAC GTC CTC
CTC AAG ATG CTC GCT TAG .

18. A polynucleotide sequence as claimed in claim 16 or 17 characterised in that it lacks some bases or includes other bases or has some of the listed codons replaced by other codons.

19. A recombinant DNA molecule characterised in that it contains a polynucleotide sequence as claimed in claim 15.

20. A recombinant DNA molecule characterised in that it contains a polynucleotide sequence as claimed in claim 16 or 17.

21. A recombinant DNA molecule as claimed in claim 20 characterised in that the polynucleotide sequence is combined with pBR322 or pUC9.

Fig.1.

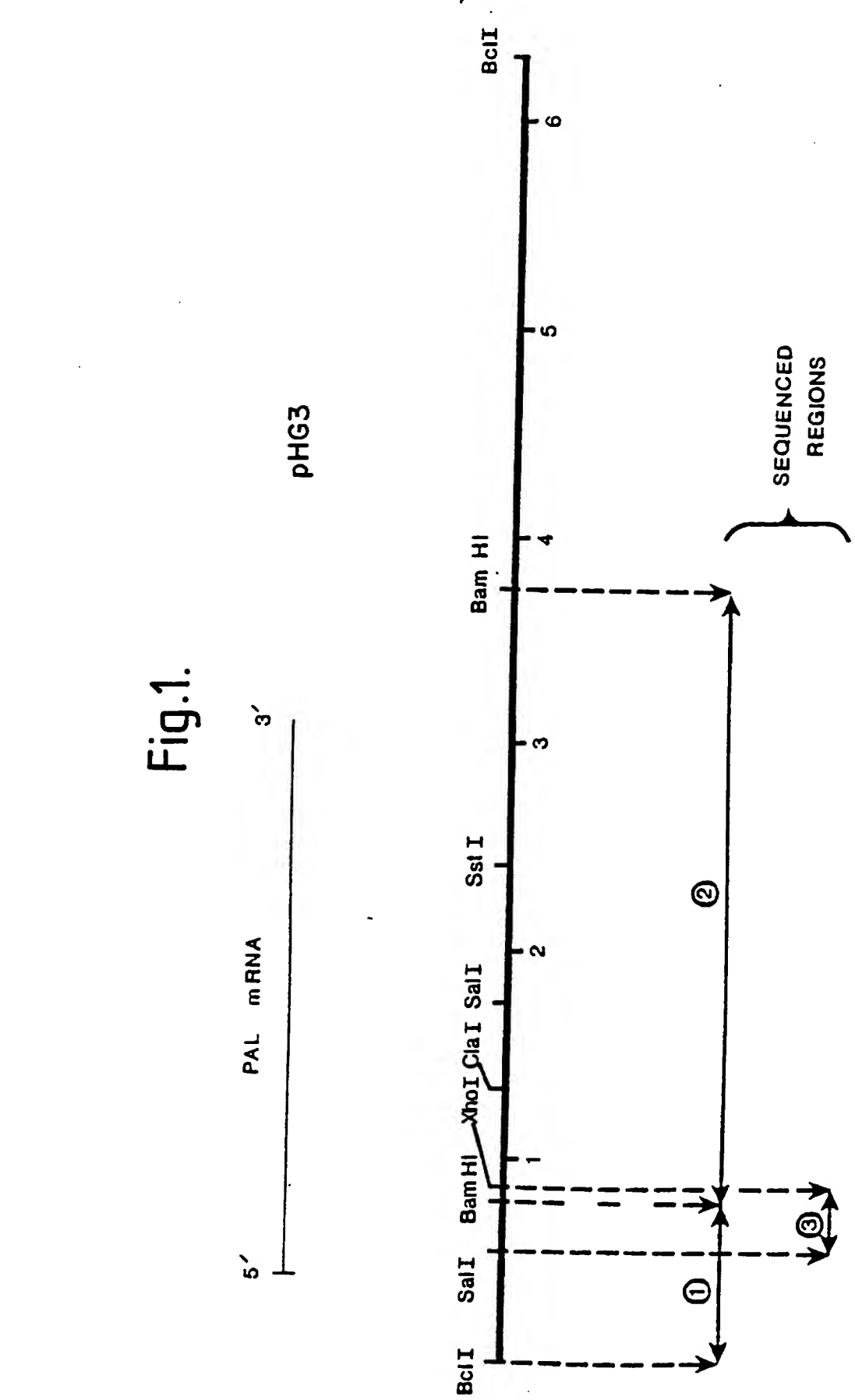
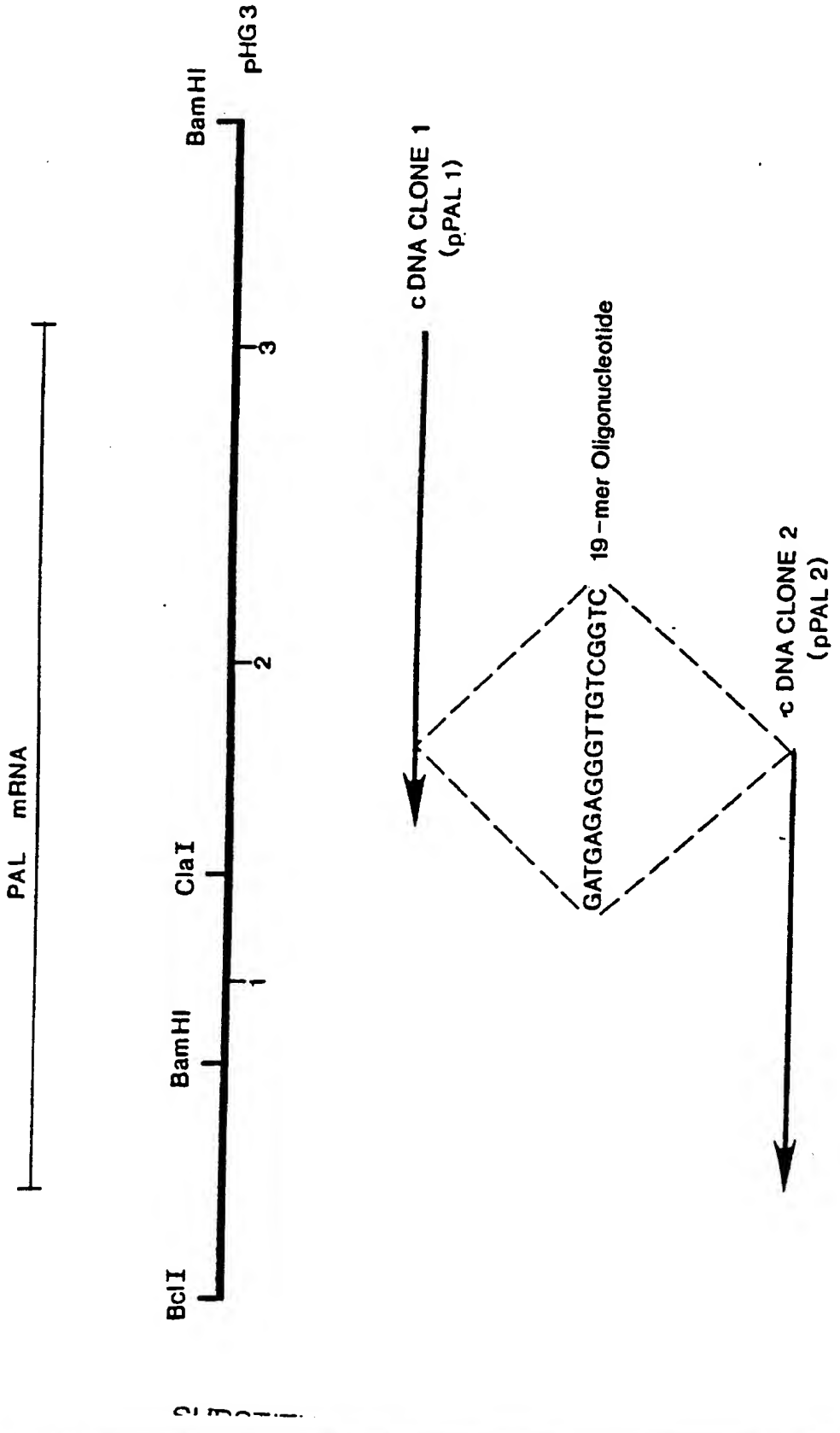
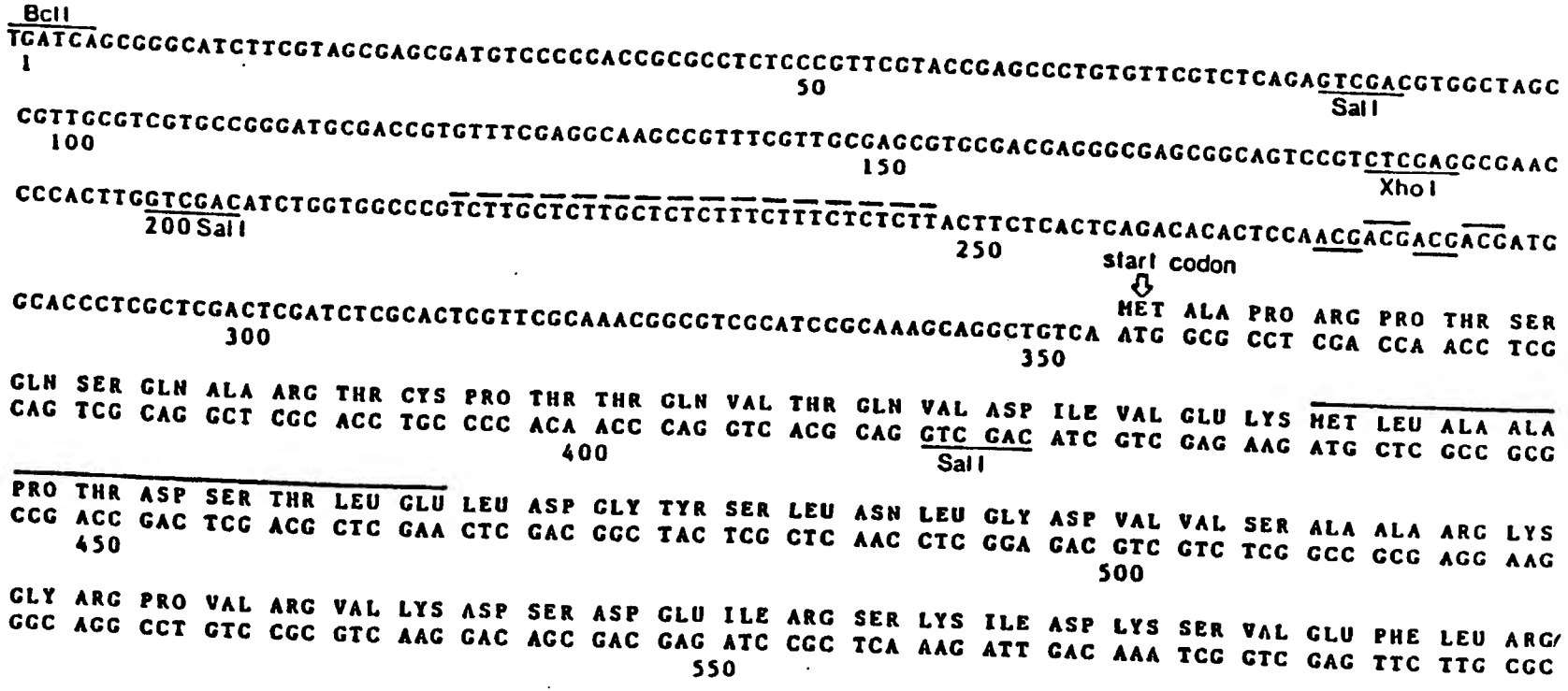

pHG3
PAL mRNA
5′
3′
BclI
SalI
BamHI
XhoI
ClaI
SalI
SstI
Bam HI
BclI
1
2
3
4
5
6
①
②
③
SEQUENCED REGIONS

Fig. 2.


PAL mRNA
pHG3
BamHI
3
2
ClaI
1
BamHI
BclI
c DNA CLONE 1
(pPAL 1)
19-mer Oligonucleotide
GATGAGAGGGTTGTCGGTC
c DNA CLONE 2
(pPAL 2)

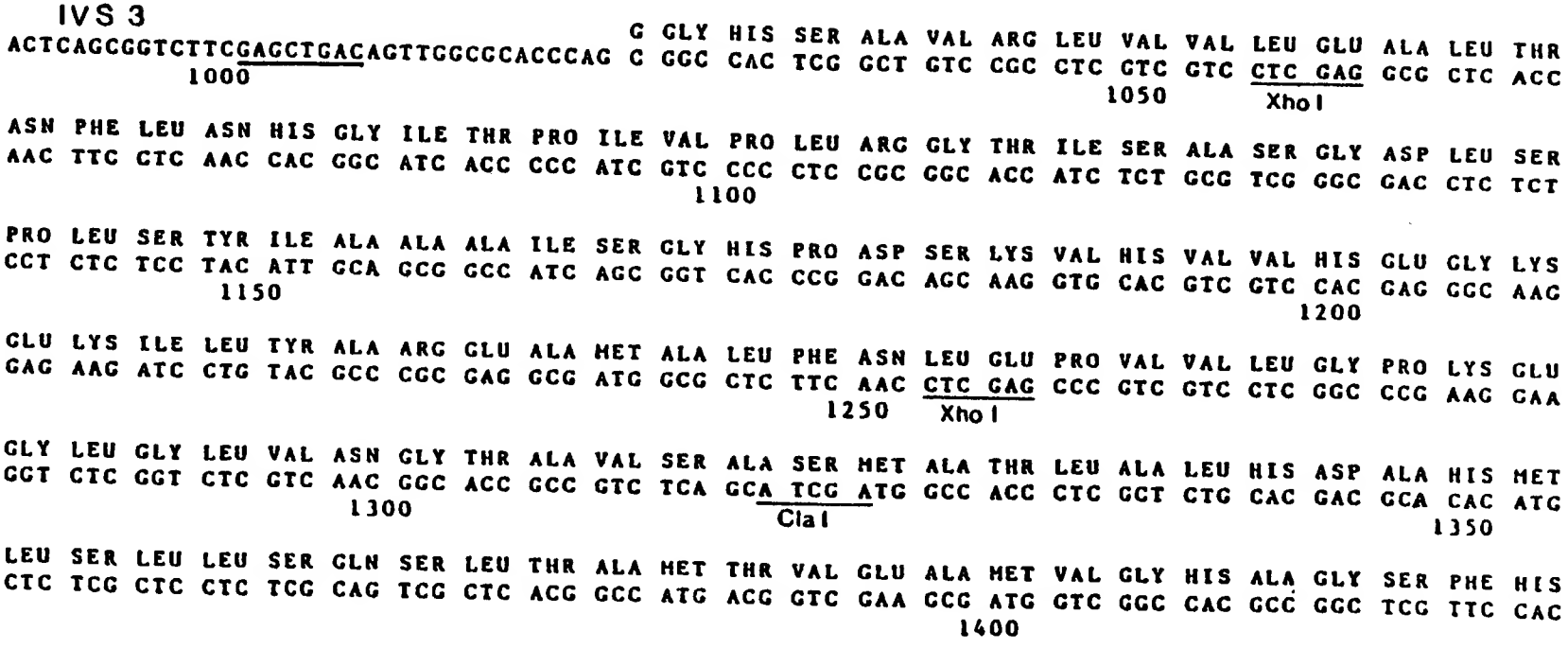
# Fig. 3A.


BclI
TGATCAGCGGGCATCTTCGTAGCGAGCGATGTCCCCCACCGCGCCTCTCCCGTTCGTACCGAGCCCTGTGTTCGTCTCAGAGTCGACGTGGCTAGC
1
50
SalI
CGTTGCGTCGTGCCCGGGATGCGACCGTGTTTCGAGGCAAGCCGTTTCGTTGCGAGCGTGCGACGAGGGCGAGCGGCAGTCCGTCTCGAGGCCGAAC
100
150
XhoI
CCCACTTGGTCGACATCTGGTGGCCCGTCTTGCTCTTGCTCTCTTTCTTTCTCTCTTACTTCTCACTCAGACACACTCCAACGACGACGACGACATG
200 SalI
250
start codon
GCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAGCAGGCTGTCA ATG GCG CCT CGA CCA ACC TCG
300
350
MET ALA PRO ARG PRO THR SER
GLN SER GLN ALA ARG THR CYS PRO THR THR GLN VAL THR GLN VAL ASP ILE VAL GLU LYS MET LEU ALA ALA
CAG TCG CAG GCT CGC ACC TGC CCC ACA ACC CAG GTC ACG CAG GTC GAC ATC GTC GAG AAG ATG CTC GCC GCG
400
SalI
PRO THR ASP SER THR LEU GLU LEU ASP GLY TYR SER LEU ASN LEU GLY ASP VAL VAL SER ALA ALA ARG LYS
CCG ACC GAC TCG ACG CTC GAA CTC GAC GGC TAC TCG CTC AAC CTC GGA GAC GTC GTC TCG GCC GCG AGG AAG
450
500
GLY ARG PRO VAL ARG VAL LYS ASP SER ASP GLU ILE ARG SER LYS ILE ASP LYS SER VAL GLU PHE LEU ARG/
GGC AGG CCT GTC CGC GTC AAG GAC AGC GAC GAG ATC CGC TCA AAG ATT GAC AAA TCG GTC GAG TTC TTG CGC
550

# Fig. 3B.

```
SER GL                                                IVS 1
TCG CA GTGAGTCGTGCTTTCGTTCTCTGGCGTCGAGAGGGCGGCACCTTCCCAAGTTGCCAAGGGACTGACTGTCGCTCTCCTGTGTCGCGCA
          600                                                                650

  N LEU SER MET SER VAL TYR GLY VAL THR THR GLY PHE GLY GLY SER ALA ASP THR ARG THR GLU ASP ALA
C A CTC TCC ATG AGC GTC TAC GGC GTC ACG ACT GGA TTT GGC GGA TCC GCA GAC ACC CGC ACC GAG GAC GCC
                        700                                 BamHI                          750

ILE SER LEU GLN LYS AL                                IVS 2                          A LEU LEU
ATC TCG CTC CAG AAG GC GTGCGTCCTCCTCCGTCTCCCTCTCGCTTCTCGACCTTCGGACTGACCGTCTTCCCGCACAG T CTC CTC
                                                    800XhoI                            XhoI

GLU HIS GLN LEU CYS GLY VAL LEU PRO SER SER PHE ASP SER PHE ARG LEU GLY ARG GLY LEU GLU ASN SER
GAG CAC CAG CTC TGC GGT GTT CTC CCT TCG TCG TTC GAC TCG TTC CGC CTC GGC CGC GGT CTC GAG AAC TCG
              850                                                          900XhoI

LEU PRO LEU GLU VAL VAL ARG GLY ALA MET THR ILE ARG VAL ASN SER LEU THR AR
CTT CCC CTC GAG GTT GTT CGC GGC GCC ATG ACA ATC CGC GTC AAC AGC TTG ACC CG GTGAGTTGCCCTCCTTACTC
         XhoI                                  950
```

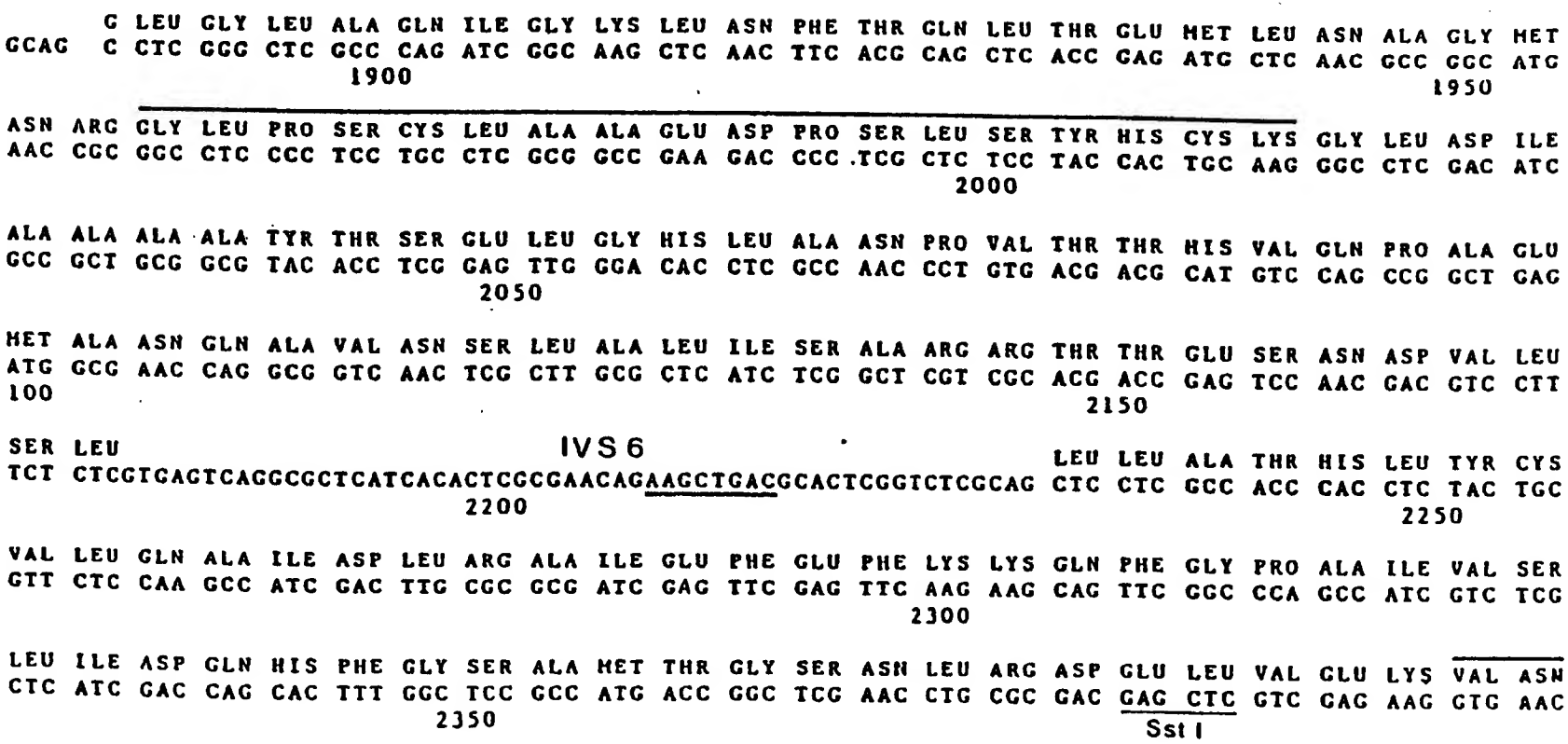
# Fig.3C.


IVS 3
G GLY HIS SER ALA VAL ARG LEU VAL VAL LEU GLU ALA LEU THR
ACTCAGCGGTCTTCGAGCTGACAGTTGGCCCACCCAG C GGC CAC TCG GCT GTC CGC CTC GTC GTC CTC GAG GCG CTC ACC
1000
1050
XhoI
ASN PHE LEU ASN HIS GLY ILE THR PRO ILE VAL PRO LEU ARG GLY THR ILE SER ALA SER GLY ASP LEU SER
AAC TTC CTC AAC CAC GGC ATC ACC CCC ATC GTC CCC CTC CGC GGC ACC ATC TCT GCG TCG GGC GAC CTC TCT
1100
PRO LEU SER TYR ILE ALA ALA ALA ILE SER GLY HIS PRO ASP SER LYS VAL HIS VAL VAL HIS GLU GLY LYS
CCT CTC TCC TAC ATT GCA GCG GCC ATC AGC GGT CAC CCG GAC AGC AAG GTG CAC GTC GTC CAC GAG GGC AAG
1150
1200
GLU LYS ILE LEU TYR ALA ARG GLU ALA MET ALA LEU PHE ASN LEU GLU PRO VAL VAL LEU GLY PRO LYS GLU
GAG AAG ATC CTG TAC GCC CGC GAG GCG ATG GCG CTC TTC AAC CTC GAG CCC GTC GTC CTC GGC CCG AAG GAA
1250
XhoI
GLY LEU GLY LEU VAL ASN GLY THR ALA VAL SER ALA SER MET ALA THR LEU ALA LEU HIS ASP ALA HIS MET
GGT CTC GGT CTC GTC AAC GGC ACC GCC GTC TCA GCA TCG ATG GCC ACC CTC GCT CTG CAC GAC GCA CAC ATG
1300
ClaI
1350
LEU SER LEU LEU SER GLN SER LEU THR ALA MET THR VAL GLU ALA MET VAL GLY HIS ALA GLY SER PHE HIS
CTC TCG CTC CTC TCG CAG TCG CTC ACG GCC ATG ACG GTC GAA GCG ATG GTC GGC CAC GCC GGC TCG TTC CAC
1400

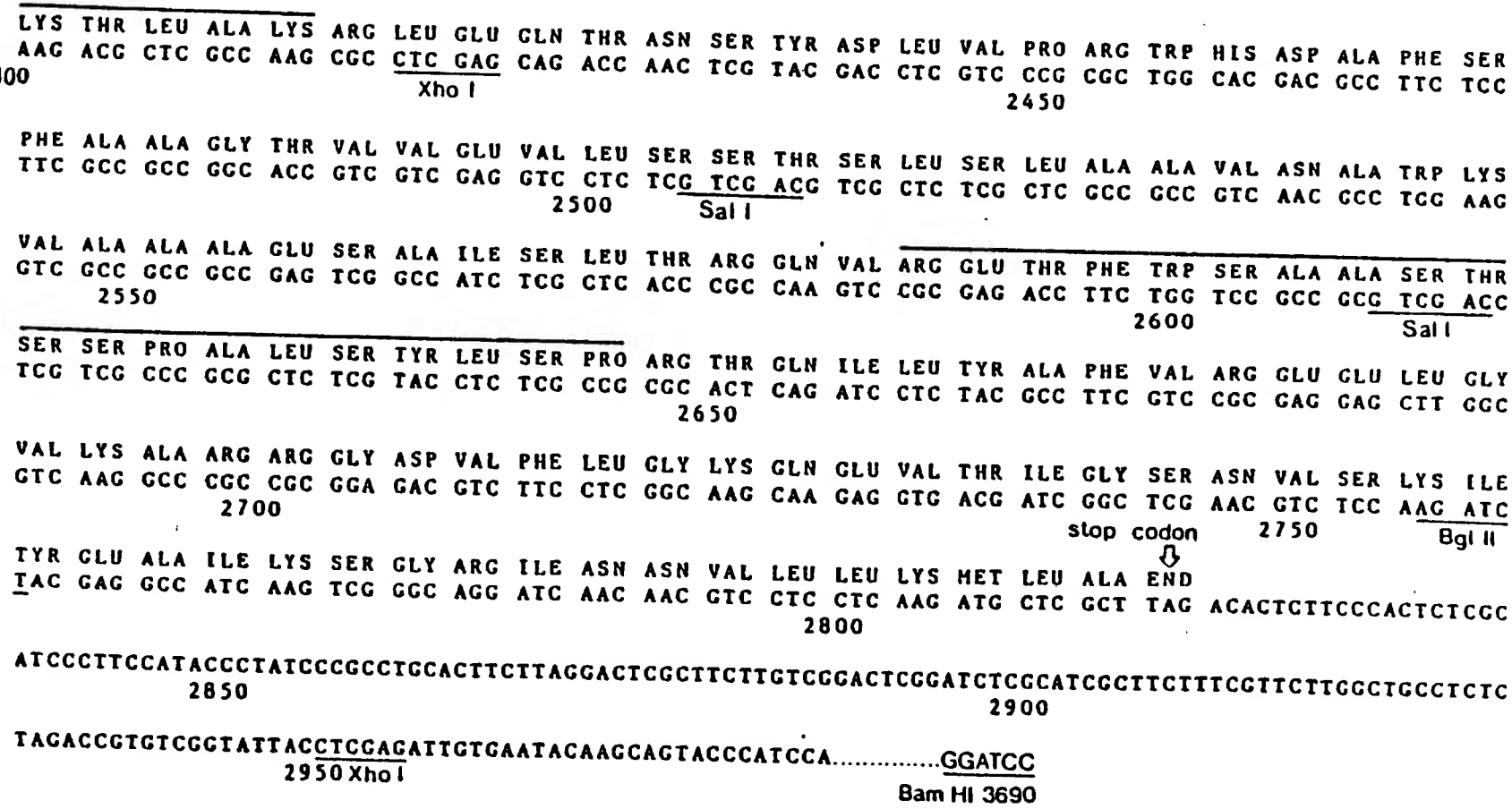
# Fig.3D.

```
PRO PHE LEU HIS ASP VAL THR ARG PRO HIS PRO THR GLN ILE GLU VAL ALA GLY ASN ILE ARG LYS LEU LEU
CCC TTC CTT CAC GAC GTC ACG CGC CCT CAC CCG ACG CAG ATC GAA GTC GCG GGA AAC ATC CGC AAG CTC CTC
                                1450                                                    Xho I
GLU GLY SER ARG PHE ALA VAL HIS HIS GLU GLU GLU VAL LYS VAL LYS ASP ASP GLU GLY ILE LEU ARG GLN
GAG GGA AGC CGC TTT GCT GTC CAC CAT GAG GAG GAG GTC AAG GTC AAG GAC GAC GAG GGC ATT CTC CGC CAG
1500                                                            1550

ASP ARG TYR PRO LEU ARG THR SER PRO GLN                              IVS 4
GAC CGC TAC CCC TTG CGC ACG TCT CCT CAG GTGCGCTTACTTCTGTTTGTTCTGCCGAAGACATGACGCTGACGTCCGCTTACTC
                  FspI                  1600                                          1650

        TRP LEU GLY PRO LEU VAL SER ASP LEU ILE HIS ALA HIS ALA VAL LEU THR ILE GLU ALA GLY GLN
GCGCAG  TGG CTC GGC CCG CTC GTC AGC GAC CTC ATT CAC GCC CAC GCC GTC CTC ACC ATC GAG GCC GGC CAG
                                                1700

SER THR THR ASP ASN PRO LEU ILE ASP VAL GLU ASN LYS THR SER HIS HIS GLY GLY ASN PHE GLN ALA ALA
TCG ACG ACC GAC AAC CCT CTC ATC GAC GTC GAG AAC AAG ACT TCG CAC CAC GGC GGC AAT TTC CAG GCT GCC
SalI     ← cDNA PRIMER    1750

ALA VAL ALA ASN THR MET GLU LYS THR AR                               IVS 5
GCT GTG GCC AAC ACC ATG GAG AAG ACT CG  GTGCGCCGCTTCACTGTGACCTGTTCTCTTGGTCTCGTGCTGACGCAGTACGCTGT
1800                                                            1850
```

# Fig.3E.


G LEU GLY LEU ALA GLN ILE GLY LYS LEU ASN PHE THR GLN LEU THR GLU MET LEU ASN ALA GLY MET
GCAG C CTC GGG CTC GCC CAG ATC GGC AAG CTC AAC TTC ACG CAG CTC ACC GAG ATG CTC AAC GCC GGC ATG
1900
1950
ASN ARG GLY LEU PRO SER CYS LEU ALA ALA GLU ASP PRO SER LEU SER TYR HIS CYS LYS GLY LEU ASP ILE
AAC CGC GGC CTC CCC TCC TGC CTC GCG GCC GAA GAC CCC TCG CTC TCC TAC CAC TGC AAG GGC CTC GAC ATC
2000
ALA ALA ALA ALA TYR THR SER GLU LEU GLY HIS LEU ALA ASN PRO VAL THR THR HIS VAL GLN PRO ALA GLU
GCC GCT GCG GCG TAC ACC TCG GAG TTG GGA CAC CTC GCC AAC CCT GTG ACG ACG CAT GTC CAG CCG GCT GAG
2050
MET ALA ASN GLN ALA VAL ASN SER LEU ALA LEU ILE SER ALA ARG ARG THR THR GLU SER ASN ASP VAL LEU
ATG GCG AAC CAG GCG GTC AAC TCG CTT GCG CTC ATC TCG GCT CGT CGC ACG ACC GAG TCC AAC GAC GTC CTT
100
2150
SER LEU IVS 6 LEU LEU ALA THR HIS LEU TYR CYS
TCT CTCGTCAGTCAGGCGCTCATCACACTCGCGAACAGAAGCTGACGCACTCGGTCTCGCAG CTC CTC GCC ACC CAC CTC TAC TGC
2200
2250
VAL LEU GLN ALA ILE ASP LEU ARG ALA ILE GLU PHE GLU PHE LYS LYS GLN PHE GLY PRO ALA ILE VAL SER
GTT CTC CAA GCC ATC GAC TTG CGC GCG ATC GAG TTC GAG TTC AAG AAG CAG TTC GGC CCA GCC ATC GTC TCG
2300
LEU ILE ASP GLN HIS PHE GLY SER ALA MET THR GLY SER ASN LEU ARG ASP GLU LEU VAL GLU LYS VAL ASN
CTC ATC GAC CAG CAC TTT GGC TCC GCC ATG ACC GGC TCG AAC CTG CGC GAC GAG CTC GTC GAG AAG GTG AAC
2350
Sst I

# Fig. 3F.


LYS THR LEU ALA LYS ARG LEU GLU GLN THR ASN SER TYR ASP LEU VAL PRO ARG TRP HIS ASP ALA PHE SER
AAG ACG CTC GCC AAG CGC CTC GAG CAG ACC AAC TCG TAC GAC CTC GTC CCG CGC TGG CAC GAC GCC TTC TCC
2400
Xho I
2450
PHE ALA ALA GLY THR VAL VAL GLU VAL LEU SER SER THR SER LEU SER LEU ALA ALA VAL ASN ALA TRP LYS
TTC GCC GCC GGC ACC GTC GTC GAG GTC CTC TCG TCG ACG TCG CTC TCG CTC GCC GCC GTC AAC GCC TGG AAG
2500
Sal I
VAL ALA ALA ALA GLU SER ALA ILE SER LEU THR ARG GLN VAL ARG GLU THR PHE TRP SER ALA ALA SER THR
GTC GCC GCC GCC GAG TCG GCC ATC TCG CTC ACC CGC CAA GTC CGC GAG ACC TTC TGG TCC GCC GCG TCG ACC
2550
2600
Sal I
SER SER PRO ALA LEU SER TYR LEU SER PRO ARG THR GLN ILE LEU TYR ALA PHE VAL ARG GLU GLU LEU GLY
TCG TCG CCC GCG CTC TCG TAC CTC TCG CCG CGC ACT CAG ATC CTC TAC GCC TTC GTC CGC GAG GAG CTT GGC
2650
VAL LYS ALA ARG ARG GLY ASP VAL PHE LEU GLY LYS GLN GLU VAL THR ILE GLY SER ASN VAL SER LYS ILE
GTC AAG GCC CGC CGC GGA GAC GTC TTC CTC GGC AAG CAA GAG GTG ACG ATC GGC TCG AAC GTC TCC AAG ATC
2700
stop codon
2750
Bgl II
TYR GLU ALA ILE LYS SER GLY ARG ILE ASN ASN VAL LEU LEU LYS MET LEU ALA END
TAC GAG GCC ATC AAG TCG GGC AGG ATC AAC AAC GTC CTC CTC AAG ATG CTC GCT TAG ACACTGTTCCCACTCTCGC
2800
ATCCCTTCCATACCCTATCCCGCCTGCACTTCTTAGGACTCGCTTCTTGTCGGACTCGGATCTCGCATCGCTTCTTTCGTTCTTGGCTGCCTCTC
2850
2900
TAGACCGTGTCGGTATTACCTCGAGATTGTGAATACAAGCAGTACCCATCCA.....GGATCC
2950 Xho I
Bam HI 3690

8/14

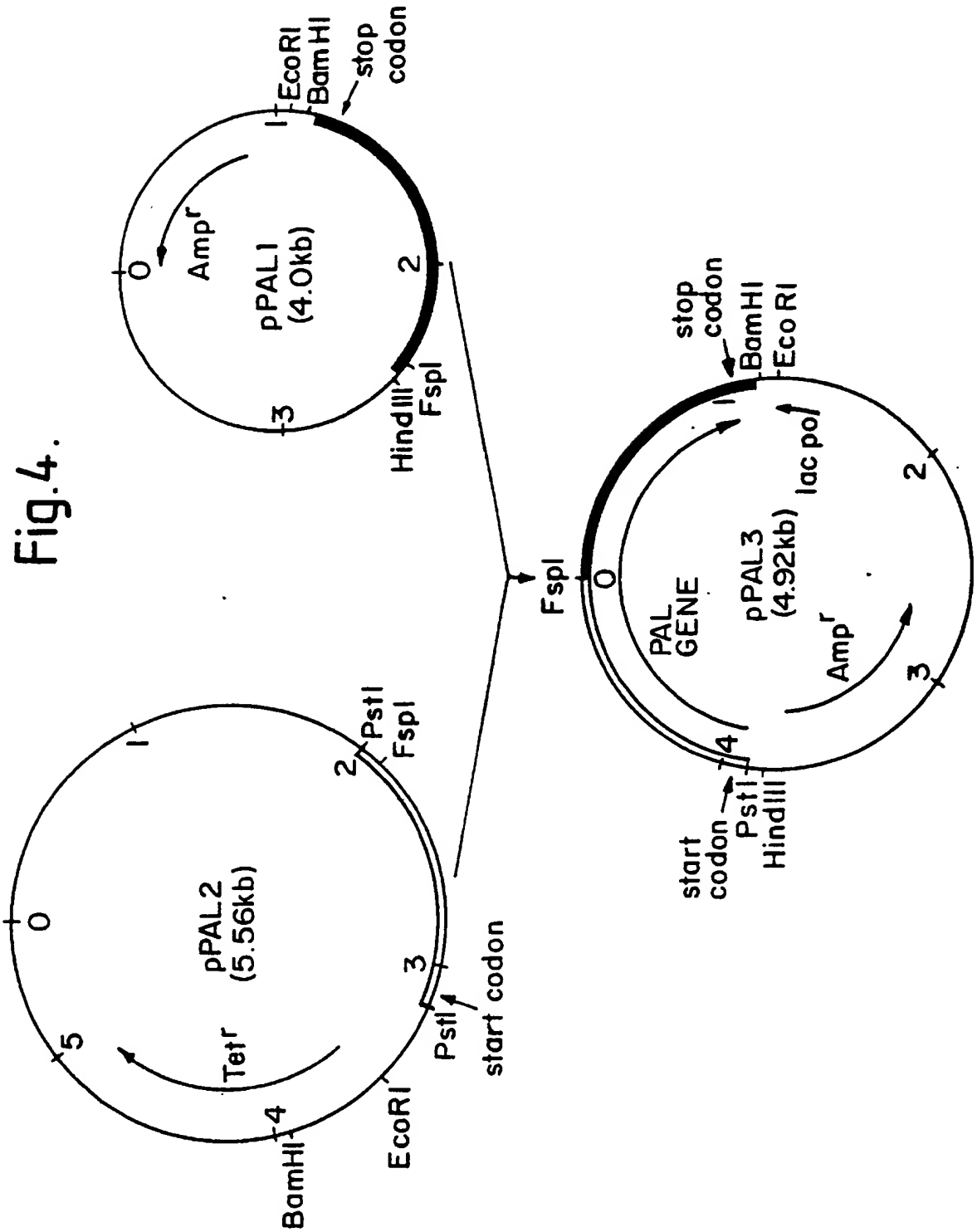

pPAL1
(4.0kb)
Amp^r
EcoRI
BamHI
stop
codon
HindIII
FspI
pPAL2
(5.56kb)
Tet^r
PstI
FspI
PstI
start codon
EcoRI
BamHI
FspI
pPAL3
(4.92kb)
PAL
GENE
Amp^r
lac pol
stop
codon
BamHI
Eco RI
start
codon
PstI
HindIII

Fig.4.

# Fig.5.

pPAL1

AAGCCT--(G)$_{15}$--
HindIII

HIS HIS GLU GLU GLU VAL LYS VAL LYS ASP ASP GLU GLY
CAC CAT GAG GAG GAG GTC AAG GTC AAG GAC GAC GAG GGC

ILE LEU ARG GLN ASP ARG TYR PRO LEU ARG THR SER PRO GLN TRP LEU GLY PRO LEU VAL SER ASP LEU ILE
ATT CTC CGC CAG GAC CGC TAC CCC TTG CGC ACG TCT CCT CAG TGG CTC GGC CCG CTC GTC AGC GAC CTC ATT

HIS ALA HIS ALA VAL LEU THR ILE GLU ALA GLY GLN SER THR THR ASP ASN PRO LEU ILE ASP VAL GLU ASN
CAC GCC CAC GCC GTC CTC ACC ATC GAG GCC GGC CAG TCG ACG ACC GAC AAC CCT CTC ATC GAC GTC GAG AAC

LYS THR SER HIS HIS GLY GLY ASN PHE GLN ALA ALA ALA VAL ALA ASN THR MET GLU LYS THR ARG LEU GLY
AAG ACT TCG CAC CAC GGC GGC AAT TTC CAG GCT GCe GCT GTG GCC AAC ACC ATG GAG AAG ACT CGC CTC GGC

LEU ALA GLN ILE GLY LYS LEU ASN PHE THR GLN LEU THR GLU MET LEU ASN ALA GLY MET ASN ARG GLY LEU
CTC GCC CAG ATC GGC AAG CTC AAC TTC ACG CAG CTC ACC GAG ATG CTC AAC GCC GGC ATG AAC CGC GGC CTC

PRO SER CYS LEU ALA ALA GLU ASP PRO SER LEU SER TYR HIS CYS LYS GLY LEU ASP ILE ALA ALA ALA ALA
CCC TCC TGC CTC GCG GCC GAA GAC CCC TCG CTC TCC TAC CAC TGC AAG GGC CTC GAC ATC GCC GCT GCG GCG

TYR THR SER GLU LEU GLY HIS LEU ALA ASN PRO VAL THR THR HIS VAL GLN PRO ALA GLU MET ALA ASN GLN
TAC ACC TCG GAG TTG GGA CAC CTC GCC AAC CCT GTG ACG ACG CAT GTC CAG CCG GCT GAG ATG GCC AAC CAG

ALA VAL ASN SER LEU ALA LEU ILE SER ALA ARG ARG THR THR GLU SER ASN ASP VAL LEU SER LEU LEU LEU
GCG GTC AAC TCG CTT GCG CTC ATC TCC GCT CGT CGC ACG ACC GAG TCC AAC GAC GTC CTT TCT CTC CTC CTC

# Fig. 5 (cont.)

```
ALA THR HIS LEU TYR CYS VAL LEU GLN ALA ILE ASP LEU ARG ALA ILE GLU PHE GLU PHE LYS LYS GLN PHE
GCC ACC CAC CTC TAC TGC GTT CTC CAA GCC ATC GAC TTG CGC GCG ATC GAG TTC GAG TTC AAG AAG CAG TTC

GLY PRO ALA ILE VAL SER LEU ILE ASP GLN HIS PHE GLY SER ALA MET THR GLY SER ASN LEU ARG ASP GLU
GGC CCA GCC ATC GTC TCG CTC ATC GAC CAG CAC TTT GGC TCC GCC ATG ACC GGC TCG AAC CTG CGC GAC GAG

LEU VAL GLU LYS VAL ASN LYS THR LEU ALA LYS ARG LEU GLU GLN THR ASN SER TYR ASP LEU VAL PRO ARG
CTC GTC GAG AAG GTG AAC AAG ACG CTC GCC AAG CGC CTC GAG CAG ACC AAC TCG TAC GAC CTC GTC CCG CGC

TRP HIS ASP ALA PHE SER PHE ALA ALA GLY THR VAL VAL GLU VAL LEU SER SER THR SER LEU SER LEU ALA
TGG CAC GAC GCC TTC TCC TTC GCC GCC GGC ACC GTC GTC GAG GTC CTC TCG TCG ACG TCG CTC TCG CTC GCC

ALA VAL ASN ALA TRP LYS VAL ALA ALA ALA GLU SER ALA ILE SER LEU THR ARG GLN VAL ARG GLU THR PHE
GCC GTC AAC GCC TGG AAG GTC GCC GCC GCC GAG TCG GCC ATC TCG CTC ACC CGC CAA GTC CGC GAG ACC TTC

TRP SER ALA ALA SER THR SER SER PRO ALA LEU SER TYR LEU SER PRO ARG THR GLN ILE LEU TYR ALA PHE
TGG TCC GCC GCG TCG ACC TCG TCG CCC GCG CTC TCG TAC CTC TCG CCG CGC ACT CAG ATC CTC TAC GCC TTC

VAL ARG GLU GLU LEU GLY VAL LYS ALA ARG ARG GLY ASP VAL PHE LEU GLY LYS GLN GLU VAL THR ILE GLY
GTC CGC GAG GAG CTT GGC GTC AAG GCC CGC CGC GGA GAC GTC TTC CTC GGC AAG CAA GAG GTG ACG ATC GGC

SER ASN VAL SER LYS ILE TYR GLU ALA ILE LYS SER GLY ARG ILE ASN ASN VAL LEU LEU LYS MET LEU ALA
TCG AAC GTC TCC AAG ATC TAC GAG GCC ATC AAG TCG GGC AGG ATC AAC AAC GTC CTC CTC AAG ATG CTC GCT

END
TAG ACACTCTTCCCACTCTCGCATCCCTTCCATACCCTATCCCGCCTGCACTTCTTAGCACTCGCTTCTTGTCGCACTCGCATCTCCCATCGCT

TCTTTCGTTCTTCGGCTGCCTCTCTACACCGTGTCGGTATTACCTCGAGATTGTCAATACAAGCAGTACCCATCCA--(A)--GGATCC
                                                                                  40   BamHI
```

# Fig. 6.

pPAL2 — *Pst I* (CTGCAG) -- (G)$_{20}$ -- CGCAAACGGCGTCGCATCCCCAAAGCAGGCTGTCA ATG GCG CCT CGA CCA ACC TCC

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | | | MET | ALA | PRO | ARG | PRO | THR | SER |
| GLN | SER | GLN | ALA | ARG | THR | CYS | PRO | THR | THR | GLN | VAL | THR | GLN | VAL | ASP | ILE | VAL | GLU | LYS | MET | LEU | ALA | ALA |
| CAG | TCG | CAG | GCT | CGC | ACC | TGC | CCC | ACA | ACC | CAG | GTC | ACG | CAG | GTC | GAC | ATC | GTC | GAG | AAG | ATG | CTC | GCC | GCG |
| PRO | THR | ASP | SER | THR | LEU | GLU | LEU | ASP | GLY | TYR | SER | LEU | ASN | LEU | GLY | ASP | VAL | VAL | SER | ALA | ALA | ARG | LYS |
| CCG | ACC | GAC | TCG | ACG | CTC | GAA | CTC | GAC | GGC | TAC | TCG | CTC | AAC | CTC | GGA | GAC | GTC | GTC | TCG | GCC | GCG | AGG | AAG |
| GLY | ARG | PRO | VAL | ARG | VAL | LYS | ASP | SER | ASP | GLU | ILE | ARG | SER | LYS | ILE | ASP | LYS | SER | VAL | GLU | PHE | LEU | ARG |
| GGC | AGG | CCT | GTC | CGC | GTC | AAG | GAC | AGC | GAC | GAG | ATC | CGC | TCA | AAG | ATT | GAC | AAA | TCG | GTC | GAG | TTC | TTG | CGC |
| SER | GLN | LEU | SER | MET | SER | VAL | TYR | GLY | VAL | THR | THR | GLY | PHE | GLY | GLY | SER | ALA | ASP | THR | ARG | THR | GLU | ASP |
| TCG | CAA | CTC | TCC | ATG | AGC | GTC | TAC | GGC | GTC | ACG | ACT | GGA | TTT | GGC | GGA | TCC | GCA | GAC | ACC | CGC | ACC | GAG | GAC |
| ALA | ILE | SER | LEU | GLN | LYS | ALA | LEU | LEU | GLU | HIS | GLN | LEU | CYS | GLY | VAL | LEU | PRO | SER | SER | PHE | ASP | SER | PHE |
| GCC | ATC | TCG | CTC | CAG | AAG | GCT | CTC | CTC | GAG | CAC | CAG | CTC | TGC | GGT | GTT | CTC | CCT | TCG | TCG | TTC | GAC | TCG | TTC |
| ARG | LEU | GLY | ARG | GLY | LEU | GLU | ASN | SER | LEU | PRO | LEU | GLU | VAL | VAL | ARG | GLY | ALA | MET | THR | ILE | ARG | VAL | ASN |
| CGC | CTC | GGC | CGC | GGT | CTC | GAG | AAC | TCG | CTT | CCC | CTC | GAG | GTT | GTT | CGC | GGC | GCC | ATG | ACA | ATC | CGC | GTC | AAC |
| SER | LEU | THR | ARG | GLY | HIS | SER | ALA | VAL | ARG | LEU | VAL | VAL | LEU | GLU | ALA | LEU | THR | ASN | PHE | LEU | ASN | HIS | GLY |
| AGC | TTG | ACC | CGC | GGC | CAC | TCG | GCT | GTC | CGC | CTC | GTC | GTC | CTC | GAG | GCG | CTC | ACC | AAC | TTC | CTC | AAC | CAC | GGC |

# Fig.6(cont.)

```
ILE THR PRO ILE VAL PRO LEU ARG GLY THR ILE SER ALA SER GLY ASP LEU SER PRO LEU SER TYR ILE ALA
ATC ACC CCC ATC GTC CCC CTC CGC GGC ACC ATC TCT GCC TCG GGC GAC CTC TCT CCT CTC TCC TAC ATT GCA

ALA ALA ILE SER GLY HIS PRO ASP SER LYS VAL HIS VAL VAL HIS GLU GLY LYS GLU LYS ILE LEU TYR ALA
GCG GCC ATC AGC GGT CAC CCG GAC AGC AAG GTG CAC GTC GTC CAC GAG GGC AAG GAG AAG ATC CTG TAC GCC

ARG GLU ALA MET ALA LEU PHE ASN LEU GLU PRO VAL VAL LEU GLY PRO LYS GLU GLY LEU GLY LEU VAL ASN
CGC CAG GCG ATG GCG CTC TTC AAC CTC GAG CCC GTC GTC CTC GGC CCG AAG GAA GGT CTC GGT CTC GTC AAC

GLY THR ALA VAL SER ALA SER MET ALA THR LEU ALA LEU HIS ASP ALA HIS MET LEU SER LEU LEU SER GLN
GGC ACC GCC GTC TCA GCA TCG ATG GCC ACC CTC GCT CTG CAC GAC GCA CAC ATC CTC TCG CTC CTC TCG CAG

SER LEU THR ALA MET THR VAL GLU ALA MET VAL GLY HIS ALA GLY SER PHE HIS PRO PHE LEU HIS ASP VAL
TCG CTC ACG GCC ATG ACG GTC GAA GCG ATG GTC GGC CAC GCC GGC TCG TTC CAC CCC TTC CTT CAC GAC GTC

THR ARG PRO HIS PRO THR GLN ILE GLU VAL ALA GLY ASN ILE ARG LYS LEU LEU GLU GLY SER ARG PHE ALA
ACG CGC CCT CAC CCG ACG CAG ATC GAA GTC GCG GGA AAC ATC CGC AAG CTC CTC GAG GGA AGC CGC TTT GCT

VAL HIS HIS GLU GLU GLU VAL LYS VAL LYS ASP ASP GLU GLY ILE LEU ARG GLN ASP ARG TYR PRO LEU ARG
GTC CAC CAT GAG GAG GAG GTC AAG GTC AAG GAC GAC GAG GGC ATT CTC CGC CAG GAC CGC TAC CCC TTG CGC

THR SER PRO GLN TRP LEU GLY PRO LEU VAL SER ASP LEU ILE HIS ALA HIS ALA VAL LEU THR ILE GLU ALA
ACG TCT CCT CAG TGG CTC GGC CCG CTC GTC AGC GAC CTC ATT CAC GCC CAC GCC GTC CTC ACC ATC GAG GCC

GLY GLN SER THR THR ASP ASN PRO LEU ILE
GGC CAG TCG ACG ACC GAC AAC CCT CTC ATC--(C)20--CTGCAG
                                          PstI
```

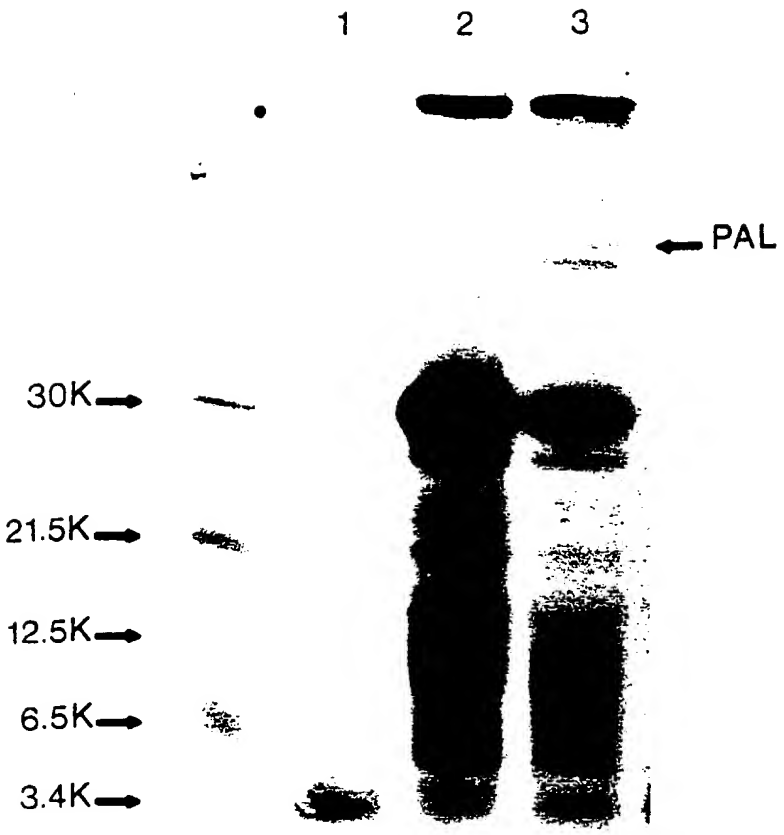



Fig.7.

# INTERNATIONAL SEARCH REPORT

International Application No PCT/GB 87/00628

## I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) [6]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC[4]: C 12 N 15/00; C 12 N 9/88

## II. FIELDS SEARCHED

Minimum Documentation Searched [7]

| Classification System | Classification Symbols |
|---|---|
| IPC[4] | C 12 N |

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [8]

## III. DOCUMENTS CONSIDERED TO BE RELEVANT [9]

| Category [10] | Citation of Document, [11] with indication, where appropriate, of the relevant passages [12] | Relevant to Claim No. [13] |
|---|---|---|
| X | EP, A, 0137280 (CETUS CORP.) 17 April 1985 see the whole document | 1 |
| Y | -- | 2-21 |
| Y | Gene, vol. 36, Elsevier Science Publishers 1985; M. Tully et al.: "Transformation of Rhodosporidium toruloides", pages 235-240, see the whole document cited in the application | 2-21 |
| | -- | |
| Y | Journal of Bacteriology, vol. 161, no. 1, January 1985, American Society for Microbiology; H.J. Gilbert et al.: "Molecular cloning of the phenylalanine ammonia lyase gene from Rhodosporidium toruloides in Escherichia coli K-12", pages 314-320, see the whole document cited in the application | 2-21 |
| | -- | ./. |

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

## IV. CERTIFICATION

| Date of the Actual Completion of the International Search | Date of Mailing of this International Search Report |
|---|---|
| 2nd December 1987 | 25 JAN 1988 |
| **International Searching Authority** | **Signature of Authorized Officer** |
| EUROPEAN PATENT OFFICE | P.C.C. VAN DER PUTTEN |

Form PCT/ISA/210 (second sheet) (January 1985)

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

| Category * | Citation of Document, with indication, where appropriate, of the relevant passages | Relevant to Claim No |
|---|---|---|
| Y | Drug Development Research, vol. 1, 1981 Alan R. Liss, Inc.; W.L. Miller et al.: "Synthesis of biologically active proteins by recombinant DNA technology", pages 435-454, see abstract; figure 1; page 422, last paragraph - page 446, paragraph 3 | 2-21 |
| X | Nucleic Acids Research, vol. 10, no. 12, 1982, IRL Press Ltd (Oxford, GB) R. Derynck et al.: "Human interferon $\gamma$ is encoded by a single class of mRNA", pages 3605-3615, see the whole document | 1 |
| Y | | 14 |
| A | Chemical Abstracts, vol. 104, 1986 (Columbus, Ohio, US) H.J. Gilbert et al.: "Cloning and expression of the Erwinia chrysanthemi asparaginase gene in Escherichia coli and Erwinia carotovora", see page 178, abstract no. 103508s, & J. Gen. Microbiol. 1986, 132(1), 151-60 | |
| A | Chemical Abstracts, vol. 97, no. 1, 5 July 1982 (Columbus, Ohio, US) H.J. Gilbert et al.: "Synthesis and degradation of phenylalanine ammonia-lyase of Rhodosporidium toruloides", see page 338, abstract no. 3335b, & J. Bacteriol. 1982, 150(2), 498-505 | |
| A | Chemical Abstracts, vol. 98, no. 17, 25 April 1983 (Columbus, Ohio, US) H.J. Gilbert et al.: "Control of synthesis of functional mRNA coding for phenylalanine ammonia-lyase from Rhodosporidium toruloides", see page 315, abstract no. 140317g, & J. Bacteriol. 1983, 153(3), 1147-54, cited in the application | |

---

# ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO. GB 8700628 SA 18566

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report. The members are as contained in the European Patent Office EDP file on 06/01/88
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| EP-A- 0137280 | 17-04-85 | AU-A- 3253084 | 07-03-85 |
| | | JP-A- 60149387 | 06-08-85 |